DOCUMENT RESUME

ED 405 365 TM 026 204

AUTHOR	Koretz, Daniel M.; And Others
TITLE	Perceived Effects of the Kentucky Instructional Results Information System (KIRIS).
INSTITUTION	Rand Corp., Santa Monica, CA. Inst. on Education and Training.
SPONS AGENCY	Ford Foundation, New York, N.Y.; Pew Charitable Trusts, Philadelphia, PA.
REPORT NO	ISBN-0-8330-2435-3; MR-792-PCT-FF
PUB DATE	96
CONTRACT	94-02248-600; 960-0402
NOTE	84p.
AVAILABLE FROM	Distribution Services, RAND, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138; fax: 310-451-6915; e-mail: order@rand.org.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Academic Achievement; Administrator Attitudes; Educational Assessment; *Educational Change; Elementary Secondary Education; *Performance Based Assessment; *Portfolio Assessment; Portfolios (Background Materials); Principals; Standards; *State Programs; Surveys; Teacher Attitudes; *Teachers; Testing Programs
IDENTIFIERS	High Stakes Tests; Impact Evaluation; *Kentucky Instructional Results Information System; *Reform Efforts

ABSTRACT

In 1994, as part of an ongoing study of the quality and effects of large-scale educational assessments, RAND began a series of studies of Kentucky's assessment, the Kentucky Instructional Results Information System (KIRIS). KIRIS is a cornerstone of Kentucky's education reform program. It exemplifies several key themes of current assessment-based reform in that it relies largely on performance assessment, measures student achievement against standards for expected performance, and is a high-stakes assessment with consequences for educators and schools. This report presents the results of part of the RAND effort, its surveys of random representative samples of fourth-gradee teachers, eighth-grade mathematics teachers, and fourth- and eighth-grade principals across Kentucky. These educators responded about their views of the program, the changes they had made in instruction, assessment, and school management in response to the program, the methods they used to prepare students for KIRIS, and their implementation of classroom-based portfolio assessment. Interviews were completed with 115 principals and 216 teachers, and mail surveys were collected from 209 teachers. Results show that teachers and principals are generally rather positive about the program, although many have reservations, and most agree that the program has caused stress. Teachers were about evenly divided over a fundamental tenet of the program: that all students can learn to a high level. Educators voiced positive and negative views of the KIRIS, but they acknowledged its impact. (Contains 20 tables and 20 references.) (SLD)

TM

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
E. D. Gill

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

ED 405 365

RAND

Perceived Effects of the Kentucky Instructional Results Information System (KIRIS)

Daniel M. Koretz, Sheila Barron,
Karen J. Mitchell, Brian M. Stecher

Institute on Education and Training

TM026204

BEST COPY AVAILABLE

The research described in this report was supported by The Pew Charitable Trusts under Grant No. 94-02248-600 and by The Ford Foundation under Grant No. 960-0402.

Library of Congress Cataloging in Publication Data

Perceived effects of the Kentucky Instructional Results Information System (KIRIS) / Daniel Koretz . . . [et al.].
p. cm.
"Supported by the Pew Charitable Trust/the Ford Foundation."
"MR-792-PCT/FF."
Includes bibliographical references (p.).
ISBN 0-8330-2435-3 (alk. paper)
1. Kentucky Instructional Results Information System.
2. School improvement programs—Kentucky. 3. Educational accountability—Kentucky. 4. Educational evaluation—Kentucky. 5. Competency based educational tests—Kentucky. 6. Education—Standards—Kentucky. 7. Educational surveys—Kentucky. I. Koretz, Daniel M. II. RAND Corporation.
LB2822.83.K4P47 1996
379.1'54'09769—dc20 96-9748
CIP

RAND is a nonprofit institution that helps improve public policy through research and analysis. RAND's publications do not necessarily reflect the opinions or policies of its research sponsors.

© Copyright 1996 RAND

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

Published 1996 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional information, contact Distribution Services: Telephone: (310) 451-7002; Fax: (310) 451-6915; Internet: order@rand.org

RAND

Perceived Effects of the Kentucky Instructional Results Information System (KIRIS)

Daniel M. Koretz, Sheila Barron, Karen J. Mitchell, Brian M. Stecher

Supported by
The Pew Charitable Trusts
The Ford Foundation

Institute on Education and Training

PREFACE

In 1994, as part of an ongoing program of research on the quality and effects of large-scale educational assessments, RAND began a series of studies of Kentucky's assessment, the Kentucky Instructional Results Information System (KIRIS). KIRIS is a cornerstone of Kentucky's education reform program (known by the acronym KERA, for Kentucky Education Reform Act). KERA is one of the most sweeping state reforms in the nation today, and KIRIS is the focus of attention nationwide. KIRIS exemplifies several key themes of current assessment-based reform. It relies largely on "performance assessment"—that is, assessment formats other than multiple choice. It measures student achievement against standards for expected performance, and those standards are intentionally set high relative to the current distribution of performance. It is a "high-stakes" assessment, although the direct consequences are for educators and schools rather than students: financial rewards for schools whose KIRIS scores improve sufficiently, and (in the near future) sanctions for schools that fail to improve.

This report presents the results of part of the RAND effort: surveys of random, representative samples of fourth-grade teachers, eighth-grade mathematics teachers, and fourth- and eighth-grade principals across Kentucky. These educators were asked for their views of the program; about the changes they had made in instruction, assessment, and school management in response to the program; about the methods they used to prepare students for KIRIS; and about their implementation of the classroom-based portfolio component of the assessment.

The surveys reported here were funded by The Pew Charitable Trusts and The Ford Foundation. The opinions presented here, however, are solely those of the authors.

This report is addressed to policymakers, educators, and educational administrators, both within Kentucky and nationwide, as well as to researchers and others interested in the issues of educational assessment and reform illustrated by KIRIS.

CONTENTS

PREFACE iii

TABLES vii

SUMMARY ix

ACKNOWLEDGMENTS xvii

1. BACKGROUND AND RESEARCH OBJECTIVES 1
   - History of KERA 1
   - Structure of the KIRIS Assessments and Accountability Index 2
   - The Content of the Surveys 3

2. PROCEDURES 4
   - Sampling 4
   - Data Collection 5
   - Data Analysis 6
   - Generalizability of Findings 7

3. SUPPORT FOR KIRIS 9
   - Overall Support for the Program 9
   - Educators' Judgments About KIRIS As a Lever for Reform 10
     - Support for Program Tenets 10
     - Perceived Usefulness for Encouraging Instructional Change 11
   - Educators' Judgments About KIRIS As a Measurement and Accountability Tool 12
     - Support for the Testing Domain, Administration, and Scoring 13
     - Perceived Accuracy of Student Achievement Information 14
     - Perceived Accuracy of School Effectiveness Data 15
     - Support for Rewards and Sanctions 17

4. IMPACT OF KIRIS ON SCHOOL MANAGEMENT AND INSTRUCTION 18
   - The Changing Role of the Principal 18
   - Expectations for Student Achievement 19
   - Course Offerings 21
   - Student Grouping and Remediation 22
   - Effects on Classroom Assessments 23
   - Effects on Instruction 23
     - Changes in Curricular Emphasis 23
       - Fourth-Grade Teachers 25
       - Eighth-Grade Mathematics Teachers 26
     - Perceived Positive and Negative Instructional Effects 26

5. PORTFOLIO PRACTICES 29
   - Training 29
   - Classroom Practices 30
   - Burdens 35
   - Scoring Criteria 36
   - Instructional Impact 37

6. PREPARING STUDENTS FOR THE KIRIS ASSESSMENTS 39
Methods for Preparing Students 40
Instructional Approaches 40
Motivational Approaches 42
Direct Test Preparation 43
Questionable Test Preparation and Administration 46
Perceived Causes of Gains on KIRIS 48

7. DISCUSSION 51
Summary of Key Findings 51
Implications 53
Lack of Support for Accountability 54
Perceived Burdens 54
Perceived Effects on Schooling 55
Ranking of Assessment Components 56
Specificity of Curriculum Frameworks 57
Effects on Equity 58
The Need to Explore the Validity of Score Gains 59
Portfolios: Impact and Validity 61
Other Issues of Validity 62
Next Steps 62

REFERENCES 65

TABLES

2.1. Sample Sizes and Response Rates for Kentucky Educators by Group and Survey Medium 5

3.1. Percentage of Teachers Reporting That Each Cognitive Component of KIRIS Has Had "a Great Deal" of Positive or Negative Effect on Instruction in Their Schools 12

3.2. Percentage of Teachers Reporting That the Information from Each Cognitive Component of KIRIS Is "Very Useful" for Improving Instruction in Their Classes 13

3.3. Percentage of Teachers Reporting That Student Achievement Information from Each KIRIS Cognitive Component Is "Somewhat" or "Very" Accurate ... 14

3.4. Percentage of Teachers Reporting Information from Each Component of KIRIS Is "Somewhat" or "Very" Reasonable for Drawing Conclusions About Educational Effectiveness 16

4.1. Percentage of Principals Reporting Rewarding Teachers Based on Their Students' KIRIS Scores 19

4.2. Percentage of all Teachers Reporting That Academic Expectations Have "Increased Greatly," by Grade and Type of Student 20

4.3. Percentage of Those Teachers Reporting a Change in Emphasis on High Standards Who Deemed That Change Harmful or Helpful 21

4.4. Percentage of Teachers Reporting Increased Use of Various Classroom Assessment Types by Grade 24

4.5. Percentage of Fourth-Grade Teachers Reporting Changes in Content Emphasis 25

4.6. Percentage of Fourth-Grade Teachers Reporting Changes in Content Emphasis Within Language Arts and Mathematics 26

4.7. Percentage of Eighth-Grade Mathematics Teachers Reporting Changes in Content Emphasis 27

5.1. Percentage of Portfolio Entries Revised 31

5.2. Important Criteria for Selecting Assessment Portfolio Entries 34

5.3. Credit for Portfolio Work in Student Grades 35

5.4. Use of Out-of-Class Portfolio-Related Preparation Time 35

6.1. Reported Hours Devoted to Test-Preparation Activities, Fourth-Grade Teachers 45

6.2. Reported Hours Devoted to Test-Preparation Activities, Eighth-Grade Mathematics Teachers 45

6.3. Percentage of Teachers Reporting Incidence of Questionable Test-Preparation and Administration Practices 48

6.4. Percentage of Teachers and Principals Reporting That Each Factor Contributed "A Great Deal" to KIRIS Gains in Their Schools 49

SUMMARY

The Kentucky Education Reform Act (KERA), arguably the most prominent state-level education reform effort in the nation, exemplifies several of the dominant themes of the current education reform movement. KERA holds schools accountable for outputs—in particular, student performance on a statewide assessment. The assessment, the Kentucky Instructional Results Information System (KIRIS), involves innovative performance-assessment components, such as portfolios. Kentucky has established performance standards on KIRIS that are high relative to the current distribution of achievement, and schools are accountable for increasing the percentage of students who reach those standards. KERA also stresses equity. For example, over the long run, all schools are held to the same performance standards and are given the same amount of time (20 years) to achieve a level equivalent to all students reaching the "proficient" standard.

KERA is an ambitious effort that requires pervasive and fundamental changes in practice, and its success is not guaranteed. The program's impact depends in large measure on the behaviors of thousands of educators statewide, and research indicates that the desired changes in practice are not easy to accomplish.

This study explored the impact of the KIRIS assessment and accountability system by surveying state-representative samples of teachers and principals in two of the grades in which KIRIS is administered. The sample included fourth-grade teachers, eighth-grade mathematics teachers, and principals whose schools included either fourth or eighth grade. All groups were surveyed in the spring of the 1994–95 school year, before the 1995 administration of KIRIS, using a computer-assisted telephone interview, and teachers were also administered a lengthy written survey. The surveys explored these educators' support for the KIRIS program, the changes they made in school management and instruction in response to it, the methods they used to prepare students for KIRIS, and teachers' implementation of the portfolio assessment program. Most questions asked about educators' experiences up to the time of the interview, but some asked specifically about the preceding (1993–94) school year. Interviews were completed with 115 principals and 216 teachers; mail surveys were collected from 209 teachers. Participation rates were excellent for principals and reasonable for teachers. The findings reported here, however, should not be taken to generalize beyond the four populations from which we sampled: elementary- and middle-school principals, fourth-grade teachers, and eighth-grade mathematics teachers.

SUPPORT FOR KIRIS

When asked a general question about their support for KIRIS, about half of the fourth-grade teachers and about 60 percent of the other three groups expressed some degree of support for the program. Few were neutral; most of the remainder expressed opposition. Respondents were about evenly split between agreement and disagreement with the basic tenet that all students can learn to high levels, and although a sizable majority consider the

short-term performance goals for their schools reasonable, few (about 15 percent) reported that they consider their long-term performance targets reasonable. Support for the accountability aspect of KIRIS was low: Only about one-fourth supported the imposition of rewards and sanctions.

The early years of the reform effort appear to have created widespread stress among educators. About three-fourths of principals reported that KIRIS places more than a minor burden on their schools, and about half of the principals and large majorities of teachers reported that the program imposes "undue pressure" on schools and students. (Some of the burdens principals reported are intended by KERA, while others are not.) About three-fourths of teachers reported that teachers' morale has declined as a result of KIRIS, and about one-third reported that students' morale had deteriorated. Virtually none reported an increase in student morale.

However, the majority of respondents reported that the program has had important payoffs. About half of the principals who reported that the program is burdensome said that the benefits outweigh the burdens, and another 16 percent said that the benefits and burdens are in balance. About three-fourths of principals reported that the program has been a useful tool for encouraging positive instructional change by reluctant teachers, and over half of teachers concurred that the program has caused resistant teachers to change. A large majority of teachers agreed that the performance-based components of KIRIS have had more than a small positive effect on instruction in their schools.

Three-fourths or more of principals agreed that KIRIS provides useful information about student achievement and reasonable information for drawing inferences about schools. Similarly, between half and four-fifths of teachers reported that the various cognitive components of KIRIS—multiple-choice items, open-response items, portfolios, and "performance events"—yielded "somewhat" or "very" accurate information about student achievement. Three-fourths of teachers agreed that KIRIS tests a wider range of skills than do multiple-choice-only tests. However, fewer than 45 percent of principals and teachers reported that KIRIS provides a better view of school effectiveness than would more conventional, commercial standardized tests. Moreover, many teachers reported negative opinions about diverse aspects of the KIRIS assessment. For example, roughly half of teachers strongly agreed that scoring standards for KIRIS are inconsistent across years and disciplines and that the curriculum content for the assessments is not defined well enough for them to prepare students adequately. Over 60 percent of principals and teachers strongly agreed that schools with highly transient populations are at an unfair disadvantage on KIRIS.

Teachers' evaluations of the effects of the four cognitive components of KIRIS conformed in part to the expectations of the program's architects but also revealed some potentially important differences among the performance-based components. Although most teachers (about 80 percent) value the information the multiple-choice items provide about student and school performance, virtually none (6 percent) reported that these items have had a great deal of positive effect on instruction. In contrast, about 40 percent of teachers reported that the open-response items and portfolios have had a great deal of positive effect.

Performance events, however, were cited as having a great deal of positive effect only about half as frequently as open-response items or portfolios—even though the primary justification for performance events is their presumed impact on instruction. In addition, open-response items (the least performance-based of the three non-multiple-choice components) were the most often cited as having had more than a small positive effect. Finally, portfolios were cited as having had negative effects on instruction almost as often as having had positive effects.

EFFECTS ON SCHOOL MANAGEMENT AND INSTRUCTION

Educators reported diverse changes in management, instruction, and classroom assessment in response to KIRIS, many of which are consistent with KERA's goals.

Principals reported widespread use of rewards and public reporting of scores to encourage teachers to improve students' scores on KIRIS. Most principals focused a great deal on encouraging teachers both to improve instruction generally and to take more focused actions (such as using test-preparation materials) to prepare students for KIRIS. About half of the middle-school principals and a third of elementary-school principals reported moving teachers from one grade to another in response to the program to place relatively more able teachers in accountability grades.

Large majorities of teachers reported making instructional changes consonant with the goals of the program. Four-fifths or more reported increasing the emphasis or instructional time devoted to problem-solving, communicating mathematics, and writing. A majority of teachers also reported increasing their own use of assessment formats other than multiple choice. Teachers reported that the portfolio program had both positive and negative effects on instruction. They said it led them to be more innovative in planning and instruction. However, portfolios also put pressure on the regular curriculum, and in response, teachers placed less emphasis on the mechanics of writing (in fourth grade) and on computation and algorithms (in eighth-grade mathematics).

A trend away from homogeneous grouping of students by ability in response to KIRIS was reported by a sizable minority of principals (about 40 percent in elementary schools and about 30 percent in middle schools). Principals reported a widespread increase in remedial services, but primarily outside of school hours. Roughly 80 percent of principals reported an increase in the number of students participating in before- or after-school remedial programs as a result of KIRIS. About half of the middle-school principals reported that KIRIS had affected course offerings in their schools: Mathematics and writing classes were the most frequently cited additions, while remedial classes and enrichment courses were the most often noted deletions.

Although KIRIS has led to an increase in teachers' expectations for most students, more teachers (24 percent) reported that expectations had increased a great deal for high-achieving students than for low-achieving (16 percent) or special-education students (12 percent). In addition, teachers were more likely to say that the increase in expectations was very helpful for high-achieving students than for low-achieving or special-education students.

PORTFOLIO PRACTICES

The responses of teachers concerning their implementation of the portfolio component of KIRIS suggest a fundamental tension between the individualization and flexibility that is desirable for instructional reform and the standardization that contributes to comparability of measurement across schools.

Although the fourth-grade writing portfolio program and the eighth-grade mathematics portfolio program are in some respects different, teachers in the two grades typically reported similar portfolio practices. In each grade, the typical teacher reported that portfolios receive 20 percent to 30 percent of class time, revision of portfolio entries is strongly encouraged, teachers provide frequent individualized assistance to students, assessment portfolio entries are selected with the scoring criteria in mind, and portfolio entries contributed about 20 percent to students' final grades.

Nonetheless, in both grades, portfolio practices varied markedly among teachers and among schools. These variations could undermine the comparability of scores across schools and therefore potentially undermine the validity of inferences about differences among schools in performance or growth.

Portfolios continue to require substantial amounts of teacher preparation time as well. Teachers reported that, in a typical month, they spend 10 hours preparing for portfolios and that between one-third and one-half of that time is devoted to scoring student work. Majorities of both groups of teachers reported that they spend too much time on scoring. The next most common preparation activities are preparing lessons and finding tasks and materials. In a heavy month, teachers devote about twice as much time to portfolio preparation as in a typical month. It is notable that 60 percent of the fourth-grade teachers and 75 percent of the eighth-grade mathematics teachers disagreed with the statement that portfolios were less of a burden in the survey year than in the preceding year.

PREPARING STUDENTS FOR KIRIS

Educators reported relying on a wide variety of approaches for preparing students for the KIRIS assessments, ranging from broad improvements in instruction to narrow and specific test preparation. Three-fourths or more of principals reported giving their teachers a great deal of encouragement to make such broad changes as raising expectations for students and focusing more on higher-order thinking skills. Almost all teachers reported focusing more than a small amount on "improving instruction generally," and more than half reported focusing a great deal on these changes.

Educators also reported widespread efforts to align instruction with KIRIS. Alignment could include both intended changes in instruction and methods that could inflate scores by focusing too specifically on the details of the assessments. Almost three-fourths of the principals reported encouraging their teachers a great deal to focus instruction on "skills or content likely to be on KIRIS," and about half reported that their schools' emphasis on such material had increased a great deal. Nearly all reported that the emphasis on such material had increased at least moderately. About 40 percent of teachers reported focusing a great deal on "increasing the match between the content of instruction and the content of

KIRIS" in their efforts to raise scores, and about half reported focusing a great deal on using "KIRIS-like tasks" in instruction. Only about one-third of principals reported that their schools' emphasis on important aspects of the pre-KIRIS curriculum had decreased somewhat, but most teachers did report a decrease in emphasis on untested material. Almost 90 percent of teachers agreed (about 40 percent strongly) KIRIS had caused teachers to "deemphasize or neglect" untested material. Half of the eighth-grade mathematics teachers indicated that they themselves emphasized some material less because of KIRIS.

Principals and teachers also reported substantial reliance on "direct test preparation": using practice tests and similar materials and providing instruction in test-taking skills. To some degree, they are encouraged to do so by the Kentucky Department of Education. About 80 percent of principals reported encouraging their teachers a great deal to use test-preparation materials, and about two-thirds said the same of instruction in test-taking skills. Almost all teachers reported focusing more than "a small amount" on test-taking skills, and about half reported focusing a great deal on them. Three-fourths reported focusing more than a small amount on practice tests and test-preparation materials, and roughly one-third reported focusing a great deal on them. Almost all teachers reported that students were given practice on the previous years' KIRIS items.

Teachers reported allocations of class times to five specific types of practice tests varied greatly. For example, the median fourth-grade teachers reported allocating about 7 instructional hours over the year to released KIRIS items, but a fourth of the teachers reported less than 3 hours, and another fourth reported 15 hours or more. The median fourth-grade teachers reported spending 15 hours on the five types of practice tests about which we asked, while a fourth of them reported spending more than 25.5 hours. The 25.5 hours reported by the teacher at the 75th percentile represents about 3 percent of total available instructional time. Eighth-grade mathematics teachers reported allocating less total time, but a larger share of their available instructional time, to the five types of practice tests. The median eighth-grade mathematics teacher reported spending about 7 hours (roughly 5 percent of instructional time) on all five types, and a fourth of the teachers reported 15 hours or more (11 percent or more of instructional time).

Appreciable minorities of teachers reported questionable test-administration practices in their schools. About one-third reported that questions are at least occasionally rephrased during testing time, and roughly one in five reported that questions about content are answered during testing, that revisions are recommended during or after testing, or that hints are provided on correct answers.

EDUCATORS' EXPLANATIONS OF THEIR GAINS ON KIRIS

Despite educators' reports of reliance on broad improvements in instruction as a method for improving scores, relatively few expressed confidence that their own schools' increases on KIRIS were largely the results of improved learning. About half of the teachers reported that increased familiarity with KIRIS and work with practice tests and test-preparation materials had contributed a great deal to their score increases, while only 16 percent said that broad improvements in knowledge and skills had contributed a great deal.

Moreover, only one-fourth reported that improvements in the knowledge and skills emphasized in KIRIS had contributed a great deal. Principals were more optimistic, but even they more often attributed gains to familiarity and test preparation than to improved knowledge and skills. However, most teachers (65 percent or more) and principals (77 percent or more) reported that improvements in knowledge and skills had contributed at least a moderate amount to their schools' gains.

IMPLICATIONS

Given KERA's scope and relative youth and the high stakes attached to KIRIS, it is not surprising that these surveys revealed a mix of positive and negative views. A reform of this scope will have unintended as well as intended effects, and it will cause some dissatisfaction even when it is working as planned. Time will be needed for educators to adapt to the program and for the Kentucky Department of Education (KDE) to make mid-course corrections. Nonetheless, the results reported here point to issues that could be addressed to improve the program's impact and suggest the need for further investigations.

Both the limited support for accountability among our respondents and principals' reports that KIRIS is burdensome may to some degree reflect positive effects of the program. For example, both principals and teachers agreed that the program is useful for inducing reluctant teachers to change their practices; one would expect some of those resistant teachers to express dissatisfaction as a result. Nonetheless, it would be a mistake to discount all of the reported concerns on those grounds. Some of the concerns expressed by our respondents are too widespread to reflect merely the views of a disgruntled minority. Moreover, in some instances, respondents pointed to specific aspects of the program that caused them concern, such as time demands and the perceived disadvantages of schools with transient populations. Centralized test-based accountability is a blunt tool, and it could prove important to explore further ways in which it is creating unintended effects.

Teachers' ratings of the four cognitive components of KIRIS could have important implications for the future design of the assessment. The component that was most often rated as having positive effects on instruction was the open-response questions, which are the least performance-based of the three non-multiple-choice components. Performance events, which involve both substantial performance elements and group work, were the least often rated of the three as having a great deal of positive impact. If these evaluations are reasonable, they suggest that it may not be necessary to rely heavily on complex performance formats, with their attendant financial costs and measurement complications, to provide an incentive for improved instruction. These findings suggest that reformers should not accept at face value the simple prescription that "good assessment is good instruction" and should instead consider the instructional effects of assessment types an open empirical question.

The fact that nearly half of the teachers strongly agreed that the curriculum frameworks are insufficiently specific is grounds for concern and additional investigation. Reformers often try to avoid making frameworks too specific to help educators focus on broader instructional goals. Inadequate specificity, however, raises the risk of inconsistent instructional change and inflated test scores (if teachers rely on the assessments themselves

as a surrogate for a curriculum framework). KDE has taken a number of steps recently to increase the specificity of curriculum frameworks; further investigation is needed to explore the adequacy and impact of those changes.

Teachers' responses to questions about expectations for students raise the prospect of negative effects on equity. Expectations are only one of many elements of equity, and teachers' perceptions in this regard may not be fully accurate. Nonetheless, the fact that teachers less often reported an increase in expectations for low-achieving students and less often reported that the change in expectations was helpful for such students is troubling. Given KERA's strong focus on equity, these results warrant further investigation, and modifications to the design or operation of KIRIS may be called for.

The findings reported here suggest that the program is meeting one of its goals—increasing the amount of student writing. At the same time, teachers' responses suggest that this change may have negative implications as well, in terms of both instructional impact and test validity. Many teachers asserted that other aspects of instruction have suffered as a result of the time students spend writing, and virtually all teachers maintained that KIRIS's emphasis on writing makes it difficult to judge the mathematical competence of some students. These concerns could be illuminated by additional research, but they also require decisions by policymakers—for example, judgments about the relative value of instruction forgone to accommodate additional writing, about the relative importance of communication compared with other aspects of mathematics, and about the trade-offs between instructional effects and test validity.

A variety of the findings reported here point to the possibility of inflated gains on KIRIS—that is, the possibility that scores have increased substantially more than mastery of the domains that the assessment is intended to represent. These findings include reports that some teachers have deemphasized or neglected untested material, reports by a large majority of teachers and principals that some schools have found ways to raise scores without improving education, reports by many teachers of the substantial reliance on test preparation and instruction on test-taking skills, many teachers' reports of substantial allocations of time to practice tests, and educators' skeptical evaluations of the causes of score gains in their own schools. The potential for inflated scores when traditional, multiple-choice tests are used for accountability is now widely accepted, and there are reasons to expect that similar problems can arise when performance assessments are used in similar ways (e.g., Koretz, forthcoming). However, these findings are in themselves not conclusive. Teachers' perceptions may not be fully accurate, and given the novelty of some of KIRIS's formats, some increase in scores stemming from familiarization could represent an increase in test validity even if it did not indicate a commensurate improvement in underlying knowledge and skills. Nonetheless, few issues are as important as whether the gains in scores on KIRIS represent real improvements in education, and the striking response patterns noted here clearly point to the need for further investigation of potential score inflation and its correlates. The finding that KIRIS gains through 1994 were not mirrored in scores in fourth-grade reading on the National Assessment of Educational Progress or scores

on the American College Testing (ACT) college-admission tests adds further urgency to this question.

Teachers' responses to questions about the portfolio component of KIRIS raise important concerns pertaining to both instructional impact and the validity of scores. Unlike the other performance-based components of KIRIS, portfolios were often cited as having both positive and negative effects on instruction. Moreover, teachers pointed to large variations in portfolio practices (e.g., differences in time allocated for revisions, assistance provided by teachers, etc.) that could undermine the validity of scores for comparisons among schools or estimates of gains. Some of the issues noted here have arisen in other portfolio programs as well (see, e.g., Koretz et al., 1994a; Stecher and Mitchell, 1995), but the high-stakes use of portfolios in Kentucky lends those issues particular importance in this instance.

Finally, a number of the results reported here suggests the need for additional validation of KIRIS, apart from the key question of possibly inflated gains. Validation, normally a complex task, is made all the more difficult in the case of KIRIS by the complex and innovative nature of the assessment itself and by the particular uses to which it is put. The task of validating KIRIS will be ongoing and will require various types of evidence. KDE has recently undertaken to increase evaluation of the validity of KIRIS, and answers to some of the concerns raised by teachers may be forthcoming over the next several years.

Taken together, the findings reported here paint a portrait of a young and complex reform that is meeting with some important initial successes but is also encountering substantial difficulties, at least in the perception of surveyed teachers and principals. KDE now has the opportunity to use these findings to guide additional inquiries and to help design programmatic changes intended to strengthen the program's impact and ameliorate unintended negative consequences.

ACKNOWLEDGMENTS

We would like to thank many people whose efforts are reflected in this report. Above all, we would like to thank the hundreds of Kentucky educators who contributed their time to participate in our surveys, especially the many teachers who took the time to complete both our telephone interview and our lengthy mail survey, despite our inability to offer them compensation for their time and effort. We also want to express our gratitude to numerous people in the Kentucky Department of Education who supported this study. In particular, we want to thank Ed Reidy, Deputy Commissioner, whose determination to use independent research as a tool for improving the Kentucky Education Reform Act made this study possible, and Brian Gong and Neal Kingston, who gave generously of their time and expertise. We also want to express our gratitude to The Pew Charitable Trusts and The Ford Foundation, which have provided the financial support for this program of research. Several colleagues at RAND assisted us. Melissa Bradley helped us convert our surveys into appropriate forms for computer-assisted telephone interviews (CATIs) and did the programming of our CATI system, and Sarah Keith assisted with numerous aspects of the study, including coding and fact-checking. Amanda Merryman of the Urban Institute assisted with coding. Barbara Thurston and Amina Assaadi prepared the document. Ed Reidy and Joan Herman reviewed the manuscript; they provided many thoughtful and helpful comments but bear no responsibility for remaining errors of omission or commission.

1. BACKGROUND AND RESEARCH OBJECTIVES

Kentucky's education reform program (the Kentucky Education Reform Act—KERA) is arguably the most prominent state educational reform in the nation. It exemplifies the currently popular focus on innovative performance assessment programs and high performance standards as mechanisms for spurring improvements in schooling. In addition, Kentucky has gone farther than most other jurisdictions in making these new assessments count: Schools now receive rewards based on improvements in their students' scores on Kentucky's assessments (the Kentucky Instructional Results Information System—KIRIS), and in the near future, they will receive sanctions if their trends in scores are sufficiently unfavorable. Numerous other reforms, including the recent reauthorization of Title I (Public Law 103-382), entail or envision high-stakes uses of performance assessment programs, and the experiences of the Kentucky program may therefore influence policy and practice in many jurisdictions.

The success of KERA will depend on the responses of educators to the KIRIS assessments and other aspects of the reform. Research on assessment-based reform suggests that school change is difficult to achieve, particularly change in teaching practices and the resulting change in student work (Druker and Shavelson, 1995). For example, early research on the reading component of the state reform program in Maryland (Guthrie et al., 1994; Afflerbach et al., 1994) suggested that change was impeded by a lack of alignment between teacher beliefs and practices and those implicit in the program, a lack of alignment between existing and mandated instruction and performance assessment, a lack of resources to help implement change, and insufficient communication from the jurisdiction about program mandates. These and other studies indicate that assessment-based reform is influenced by a number of complex school and classroom variables, including incentive systems, local beliefs and norms, financial resources, and available materials and support services (see also McLaughlin, 1990).

The study reported here investigated the impact of KERA and KIRIS on education in Kentucky by surveying representative samples of teachers and principals about their views of the program and their responses to it.

HISTORY OF KERA

KERA stems from a 1989 decision of the Kentucky Supreme Court that declared the state's school system unconstitutional. In response, the legislature passed KERA, which created "an entirely new system of public education, supported by a substantial increase in funding and a more equitable allocation across districts" (David, 1994, p. 707) and called for a high-stakes performance-based assessment system:

> The Kentucky Education Reform Act of 1990 (KERA) was a bold move by the General Assembly to establish a framework for major revision of Kentucky's educational system. KERA established goals for the educational system, provided a procedure by which those goals would be defined and assessed, and created a series of rewards and sanctions to be

> associated with performance of schools on those assessments. As a direct result of KERA, the Department of Education established the Kentucky Instructional Results Information System (KIRIS) (Advanced Systems in Measurement and Evaluation, 1993, p. 1).

KERA holds schools accountable for performance on the KIRIS assessment. As noted, schools are assigned rewards or sanctions based on changes in their performance, but neither KERA nor the implementing policies of the Kentucky Department of Education establish rewards or sanctions for individual students. (Schools may take actions that impose consequences for students for their performance on KIRIS; for example, we surveyed educators about their use of KIRIS-related work in assigning class grades and making instructional placements.)

The KIRIS assessments were first implemented in 1991–92. Cash rewards for schools showing improvement in their KIRIS scores were first awarded in 1995, reflecting the end of the first accountability cycle in 1994. Sanctions as well as rewards will be assigned at the end of the second accountability cycle, which ends in 1996 (Kentucky Department of Education, 1993, p. 3).

STRUCTURE OF THE KIRIS ASSESSMENTS AND ACCOUNTABILITY INDEX

KERA holds schools accountable for performance on an accountability index that has both cognitive components (performance on the KIRIS assessments) and noncognitive components (data on dropout rates, retention in grade, attendance, and the transition from school to work). The KIRIS assessments dominate the accountability index, accounting for five-sixths of schools' scores.

KERA establishes a biennial accountability cycle. The starting and ending points in each cycle are measured with two years of data to improve the reliability of scores. (The exception was the initial baseline, which was based on a single year of data.) Each two-year average serves both as the end point for one biennium and as the baseline for the next.

Initially, all data were to be collected in three grades: fourth, eighth, and twelfth. Starting in the 1994–95 school year, however, some testing will be done in each of five grades. (All aspects of twelfth-grade testing except portfolios have been moved to grade 11, and mathematics portfolios have been moved from fourth grade to fifth grade.) However, each type of testing (for example, writing portfolios) is still carried out in only three grades.

In the first year, the school accountability index was based on assessments in reading, writing, mathematics, and social studies. In the second year, an interdisciplinary component was added that incorporated questions in arts and humanities, practical living, and vocational education. Data from these were not included in the accountability index for the first biennium but were included in the baseline index for the second biennium (Kentucky Department of Education, 1994, p. 1-1).

KIRIS originally included four types of tasks. The "transitional assessment," so named to denote an expected transition to a fully performance-based assessment program, originally included both multiple-choice items and open-response pencil-and-paper questions. During the first biennium, a decision was reached not to count the multiple-choice items in

the accountability index and to remove them from future assessments. However, in response to external criticism, the Kentucky Department of Education has decided to reintroduce multiple-choice questions. Open-ended written tasks include short-answer and essay questions. Performance events are hybrids of group and individual activities; randomly selected groups of students work together on a task for 10 to 20 minutes and then work alone for 25 to 35 minutes to produce a written product pertaining to that task. Portfolios in both writing and mathematics are compiled over substantial periods and are scored by classroom teachers following state guidelines.

Student performance on KIRIS is scored against three standards: apprentice, proficient, and distinguished. Students who fail to reach the first standard are labeled "novice" and are assigned zero on that task or assessment. Only a small proportion of Kentucky students are exempted from KIRIS (e.g., on the basis of severe disabilities), and all nonexempt students who are not tested are assigned scores of zero.

A formula is used to convert scores on the KIRIS assessments and data on the noncognitive measures into a single school accountability index. Schools are rewarded or sanctioned on the basis of the amount of change on that index, relative to a target, or threshold, based on their initial performance. The assumption underlying the performance threshold is that each school should be able to reach an index equivalent to having all of its students at the proficient level—a value of 100—within 20 years. Hence each school's improvement threshold equals its baseline plus 10 percent of the difference between its baseline and an index of 100. (This has the effect of requiring larger changes on the index for schools with lower initial achievement.) Improvement beyond that specified by the threshold garners rewards. Sanctions are assigned to schools based on the severity of a failure to meet their improvement thresholds.

THE CONTENT OF THE SURVEYS

The surveys reported here focused primarily on the KIRIS assessments and the accountability system based on them; many important aspects of the KERA reform, such as ungraded primary education, were addressed only in passing or not at all. The surveys explored educators' support for KIRIS and the accountability system, changes they made to school organization and management in response to the program, changes in classroom instruction and assessment, implementation of the portfolio program, and methods used to prepare students for KIRIS. While many of the questions pertained to the assessment or accountability systems in their entirety, others focused on specific details of the program. For example, educators were asked specific questions about their accountability indexes and targets, and teachers were asked to comment specifically on each of the four components of the KIRIS assessment (multiple-choice items, open-response items, performance events, and portfolios).

The results of this research should be of interest to participants and stakeholders in Kentucky schools and to educators and policymakers in other states contemplating education reform based on innovative performance-assessment programs, school-level accountability for student performance, or both.

2. PROCEDURES

This report summarizes the results from mail and telephone surveys that were administered during the 1994–95 school year to assess Kentucky principals' and teachers' opinions of KIRIS and their instructional and managerial responses to it. A total of 115 principals (51 from elementary schools and 64 from middle schools) and 216 teachers (112 fourth-grade teachers and 104 eighth-grade mathematics teachers) were interviewed.

SAMPLING

A multistage design was employed for sampling purposes. Stratification was based on school size and change on the KIRIS accountability index. Within each stratum, a random sample of Kentucky's elementary and middle schools was chosen. Schools with fewer than 10 students in the grade assessed using KIRIS (fourth or eighth grade) were excluded, as were schools that could be identified by name as serving special populations. In the fourth-grade sample, a sampling rate of approximately 10 percent was then used to select at random the schools to be contacted. In the eighth-grade sample, a sampling rate of approximately 28 percent was used. Eighty elementary schools and 98 middle schools were sampled.

Each school was initially contacted, at the beginning of 1995, by means of a letter to the school's principal explaining the study and indicating that we would call the principal to request his or her participation. When subsequently contacted by telephone, principals were asked if they would be willing to participate in the survey. Willing principals were then asked a number of screening questions that determined their eligibility for the principal survey. Because the survey focused in part on changes in practice over time, we attempted to interview only principals who had served as an administrator in a Kentucky school at the same level (e.g., middle school) for at least four years, including the school year of the survey (i.e., since 1991–92). Approximately 29 percent of the principals in the elementary-school sample and 22 percent of the principals in the middle-school sample were too new to administration to satisfy this eligibility screen.

Several principals were unwilling to participate (because, for example, they felt they were too busy) and offered to allow us to contact an assistant principal or other administrator and request that person's participation. In these cases, the offered proxy was required to meet the same eligibility requirement as principals. Principals who were unwilling to participate and who did not offer a proxy were considered refusals, as were principals whom we could not contact.

All sampled principals, regardless of their own eligibility, were asked to provide names of teachers for our survey of teachers, and 90 percent of the total sample of principals (including ineligible principals) provided this information. Principals selected for the elementary-school sample were asked to provide the names of teachers who were teaching fourth-grade students. In schools with three or fewer fourth-grade teachers, all teachers were contacted and were asked to participate in the study. In schools with more than three

fourth-grade teachers, a random sample of three teachers was selected, and only those teachers were contacted and asked to participate in the study. A total of 186 fourth-grade teachers were selected for inclusion in the study.

Principals selected for the middle-school sample were asked to provide names of teachers teaching mathematics to eighth-grade students. All eighth-grade mathematics teachers in sampled schools were contacted and asked to participate in the study. A total of 175 eighth-grade mathematics teachers was selected for inclusion in the study.

Because of our focus on changes in practice, we also attempted to interview only teachers who had taught the relevant grade (and subject area in the case of eighth-grade teachers) in one of the three years (1991–92, 1992–93, or 1993–94) preceding the school year in which the survey took place. Thirteen percent of the fourth-grade teachers contacted, and 18 percent of eighth-grade mathematics teachers, did not satisfy this eligibility screen.

DATA COLLECTION

Data were collected using both a computer-assisted telephone interview (CATI) and a written survey. Both the interviews and the surveys were developed at RAND by project staff and piloted with Kentucky educators. CATIs were used to collect data from principals. Both CATIs and written surveys were used to collect data from teachers. We attempted to get both interview and written survey data from all of the eligible teachers; however, for some teachers we succeeded in obtaining data from only one of the two sources.

Participation rates were very high (over 80 percent) for the principal sample (Table 2.1).[1] The participation rates for teachers were about 70 percent in three cases (both CATIs and the fourth-grade mail survey) and 65 percent for the eighth-grade mail surveys. Because some teachers completed an interview but not a survey or vice-versa, somewhat fewer teachers (58 percent in both grades) completed both instruments.

Principal interviews were designed to collect information about school demographics, general support for the reform effort, the principal's own responses to the reform (including her or his role as an instructional leader), and effects of the reform effort on the school, its teachers, and its students. Questions were also asked about how the information provided by KIRIS is used in the school and the burdens imposed by the program.

Table 2.1

Sample Sizes and Response Rates for Kentucky Educators by Group and Survey Medium

	Fourth Grade	Eighth Grade
Principal CATI	51 (89%)	64 (84%)
Teacher CATI	112 (70%)	104 (71%)
Teacher mail survey	114 (71%)	95 (65%)
Both teacher instruments	94 (58%)	84 (58%)

[1]To be conservative in estimating participation rates, we assumed that principals and teachers we failed to contact were eligible and refused to participate.

Teacher interviews included some questions that mirrored questions asked of principals (e.g., questions about general support) as well as questions about test preparation practices, classroom practices, and understanding of the KIRIS program. The written surveys asked teachers questions concerning their opinions of the usefulness and reasonableness of KIRIS results for specific purposes, questions concerning school climate, and a number of questions concerning the portfolio program. Also, more questions were asked on the written surveys about classroom assessment practices and preparation for KIRIS.

Most questions in all of the instruments asked about educators' opinions at the time of the survey or about their experiences with KIRIS or KERA up to that time. However, a small number of questions focused specifically on the previous school year (1993–94). For example, questions about test-administration practices focused on the 1993–94 school year because KIRIS had not yet been administered in 1995 at the time of our surveys.

The vast majority of questions asked were presented in a closed format (e.g. 3-, 4-, and 5-point Likert scales and yes/no questions) or required the respondent to answer with a single number. A small number of open-ended items were also asked. An effort was made to balance questions, in that some implied a positive view of the program, while others implied a negative view or raised possible criticisms. (Many were neutral.) This balance was particularly important in the case of the relatively few questions that asked respondents to express the strength of their agreement or disagreement with assertions, such as "all students can learn to a high level" and "the emphasis on high standards for all students is putting undue pressure on schools and students."

DATA ANALYSIS

Frequencies for the Likert questions were calculated both for each grade and for the two grades combined. Questions requiring a numerical response were summarized using means, standard deviations, medians, and other selected quantiles. When data were combined across grades, they were weighted to equal counts for the two grades. Weighting was done at the instrument level—i.e., teacher CATI results were weighted separately from the results of the teacher mail surveys.

Responses to open-ended items in the CATIs were audiotaped whenever the respondent gave his or her permission for taping.[2] These responses were coded from the tapes after the survey was completed, which allowed us to use a substantial number of responses in devising a coding scheme.

Because this study was designed to provide a description of the implementation and impact of the KIRIS program, we were interested in collecting a wide variety of information from principals and teachers. Accordingly, we decided to devote resources to increasing the breadth of data collected at the expense of smaller samples and lower statistical power. Because of the large amount of information collected from each principal and teacher and the

[2]In the small number of cases in which the respondent was unwilling to be taped, the interviewer typed in the respondent's comments.

relatively small size of the samples, we have taken a descriptive rather than statistical approach throughout this report.

Although we chose not to conduct formal statistical tests, we were guided in our presentation of results by the degree of confidence we believe can be placed in specific results. For example, we were relatively confident in the estimates of the percentage of principals or teachers reporting a specific view or practice. Confidence intervals for simple random samples can provide a rough guide to the level of confidence one can have in estimates from these surveys. For example, the 95 percent confidence intervals for proportions of .60 (or .40) would be ± .098 for a simple random sample of 100 and ± .069 for a simple random sample of 200. The corresponding error bands for a proportion of .20 (or .80) would be smaller: ± .080 for a sample of 100, and ± .057 for a sample of 200. These examples are typical of the results we reported.[3] Comparisons across groups (e.g., the percentage of elementary principals as compared with the percentage of middle-school principals) generally have larger margins of error, and we therefore present such differences only when they are large or, for other reasons, are suggestive.

There are, of course, other threats to the validity of the conclusions we reach in this study. Not all educators who were sampled chose to participate, and this may introduce some degree of bias into the reported results. In addition, as in all surveys, respondents may have tended to shade their responses in the direction they considered socially desirable, and some results therefore may be less than accurate reflections of educators' true beliefs. This "social desirability bias" could include a tendency to underreport or fail to report activities that are considered of questionable propriety.

GENERALIZABILITY OF FINDINGS

The resulting teacher samples are representative of Kentucky's fourth-grade teachers and eighth-grade mathematics teachers with sufficient experience. The resulting principal samples are representative of elementary and middle-school principals, with the same restrictions. For simplicity, we sometimes refer to these groups together as "educators," but this does not imply that findings would have been similar for other groups of Kentucky educators.

We compared the schools in our sample with those in the state as a whole on school size (as estimated by the number of students tested), on the baseline KIRIS accountability index, and on change in the KIRIS accountability index during the first biennium. School-level statistics for all four groups were quite similar to the population values for the baseline index and for change in the index. For example, the median baseline score for eighth grade was 36 for all schools, 37 for the schools from which we interviewed principals, and 36 for the schools from which we interviewed at least one teacher. For fourth grade, the median baseline score for all schools was 33, the median for the principal sample was 33, and the

[3]These simple-random-sample estimates provide only a rough guide to confidence bands for the results of these surveys. The clustering (within some schools) in our samples of teachers would increase the margin of error relative to that in a simple random sample; a finite sample correction would reduce it modestly.

median for the schools in which we interviewed at least one teacher was 32. For a number of students tested, the fourth-grade statistics were quite similar to the population, whereas the eighth-grade statistics were higher than the population as a whole. That is, we obtained data from principals and teachers in middle schools that were on average somewhat larger than the population of middle schools as a whole.

Because the sum of weights was equal in the two grades, both grades counted equally when we combined results across grades. Had we instead weighted each grade to reflect its numbers in the population of teachers, fourth-grade teachers would have counted far more in our results, even though they teach essentially the same number of students as do the eighth-grade mathematics teachers. However, in most instances, results were combined only when they were similar in the two grades, so weighting proportionally would not have greatly changed the results reported here.

3. SUPPORT FOR KIRIS

We asked Kentucky educators about their support for KIRIS, both overall and with respect to each of the three functions KIRIS serves: inducing reform, monitoring school performance, and providing the basis for accountability. Additionally, we asked principals and teachers about their use of information from KIRIS in drawing conclusions about student achievement and educational effectiveness. While respondents expressed positive general views of KIRIS as a tool for inducing reform and monitoring school performance, principals and teachers reported a mix of opinions about the quality of KIRIS as an assessment, on a number of administration and scoring issues, and on factors that potentially distort KIRIS comparisons among schools and over time.

OVERALL SUPPORT FOR THE PROGRAM

When asked global questions about their support for KIRIS, the majority of principals and eighth-grade mathematics teachers (about 60 percent) said they support the program as a whole. About half of the fourth-grade teachers said the same (47 percent). With the exception of elementary-school principals (who were almost evenly split), most supporters characterized themselves as "somewhat" supportive, rather than "strongly" supportive of the program. Substantially more principals than teachers reported strong support for KIRIS. Few respondents held neutral views; over 35 percent of principals and eighth-grade mathematics teachers reported opposition to KIRIS. About half of the fourth-grade teachers (47 percent) said they somewhat or strongly oppose the program.

Support for the rest of KERA, such as site-based management and the ungraded primary program, was more widespread than support for KIRIS. About three-fourths of principals (77 percent) said they support the rest of KERA, and approximately half of the elementary-school (47 percent) and a third of middle-school principals (32 percent) characterized their support of KERA as strong—many more than reported strong support for KIRIS.

Almost half of the principals (44 percent) said their attitude toward KIRIS has become more positive over the last few years; slightly more elementary- than middle-school principals expressed that view. About a quarter of principals (26 percent), however, said their attitude has become more negative.

When asked about the burden KIRIS imposes on their school, almost three-fourths of principals (74 percent) said it imposes more than a minor burden (reporting a "moderate" or "great" burden), and substantial minorities of principals said the program imposes a great burden (25 percent and 39 percent of elementary- and middle-school principals, respectively). Virtually all principals (98 percent) agreed that time demands are an important reason they find KIRIS burdensome. Large majorities agreed that the need for staff retraining (83 percent), that staff stress or low morale (80 percent), and that the need for rapid instructional change (74 percent) are burdensome. Other cited program burdens include

unclear achievement targets (66 percent), management and record keeping (58 percent), and difficulty motivating staff (56 percent). The KERA program intentionally creates a need for staff retraining and rapid instructional change, and the fact that principals currently find those burdensome could be an indication that the program is successful in that regard. The other noted burdens, however, are not intended effects of the program.

The majority of principals who perceived the program as burdensome, however, said that the benefits of the program balance or outweigh the burdens it imposes. About half of the principals (48 percent) who said KIRIS is burdensome agreed that the program's benefits as a tool for improving the quality of instruction in their school are greater than the burdens it imposes. An additional 16 percent said that benefits and burdens balance. Moreover, 65 percent of the principals said the program has become easier to accommodate in their schools in the several years it has been in place. Very few (13 percent) said the program has become harder to accommodate.

Many principals and teachers agreed that the emphasis on high standards for all students is putting "undue pressure" on schools and students. Far more teachers than principals (56 percent) held this view. Almost all fourth-grade teachers (93 percent) and about three-fourths of eighth-grade mathematics teachers (76 percent) agreed that high standards are putting undue pressure on schools and students, and about half of the teachers at both levels (52 percent) strongly agreed.

Even more teachers reported strong views of undue pressure when asked a more specific question about the pressure they feel to improve student performance on KIRIS. Virtually all (98 percent) agreed that teachers are under undue pressure to improve student performance on KIRIS. Many more teachers (over 80 percent) strongly agreed with this statement than with any other statement about support about which we asked. Reports of undue pressure, however, could reflect either intended or unintended outcomes. In some instances, they may point to individuals who are reluctantly implementing program goals and resent the pressure to do so. In other instances, they may indicate weaknesses in the program.

EDUCATORS' JUDGMENTS ABOUT KIRIS AS A LEVER FOR REFORM

The sub-subsections below discuss educators' judgments about two functions served by KIRIS: serving as an agent of reform and as a tool for monitoring and holding schools accountable for student achievement.

Support for Program Tenets

KIRIS and KERA are based on a number of fundamental beliefs about student performance and educational opportunity. Like school reformers elsewhere, KIRIS's architects assert that all students can learn to high levels and that the job of Kentucky educators is to provide the means by which that happens. We asked teachers whether they agree with some of the premises of the program.

Teachers were divided in their opinion about one of the fundamental tenets of the program—that all students can learn to a high level. About half of the teachers (46 percent)

said they agree with this premise, and about half disagreed (54 percent). Nonetheless, the large majority of teachers (83 percent) agreed that regardless of whether or not it is possible for all students to learn to a high level, it is the right message to give to Kentucky students. However, almost none of the teachers (9 percent) agreed with the notion that all students can reach the *same* high level of performance; in fact, the vast majority of teachers (90 percent) said that novice—the level assigned to students who fail to reach the lowest KIRIS standard (apprentice)—is a high level of performance for some students.

Principals and teachers expressed mixed opinions about the program's expectations for school performance. A large majority of teachers (66 percent) and principals (70 percent) agreed that the current improvement threshold for their school is a realistic goal, but only a few (15 and 13 percent of principals and teachers, respectively) reported that they consider the long-term goal of obtaining an accountability index of 100 to be realistic.

Perceived Usefulness for Encouraging Instructional Change

Most principals and teachers were positive about KIRIS's value as an agent of reform. About three-quarters said KIRIS has been useful for encouraging positive instructional change among teachers who are very resistant to making changes to their instruction (77 percent); approximately a quarter said it has been very useful in this respect (24 percent). Over half of the teachers (57 percent) agreed that KIRIS has caused some teachers who are resistant to change to improve their instruction.

The relative benefits of KIRIS's cognitive components—multiple-choice items, open-response items, performance events, and portfolios—have been the focus of debate, so we asked teachers to provide their opinions of the instructional effects of each. Fourth-grade teachers were asked for judgments about the impact of the KIRIS components on instruction generally, while eighth-grade mathematics teachers were asked about effects on mathematics instruction.

Perhaps predictably, teachers more often credited the three performance-based components of KIRIS than the multiple-choice items as having had a positive effect on instruction, but a few of the distinctions among the cognitive components were striking. About two-thirds of teachers cited portfolios and performance events as having had more than small positive effects on instruction in their schools (reporting a "moderate amount" or "a great deal" of positive effect).[4] More teachers (about 80 percent) cited the open-response items as having had more than a small positive effect. While this difference is small in terms of statistical confidence, it is striking nonetheless, because the open-response items, which typically entail only a brief written response, are the least performance-based of the three performance components. Fewer teachers (41 percent) reported more than small positive effects for multiple-choice items.

A few important changes in this pattern appear, however, when only responses citing "a great deal" of positive instructional impact are considered. In that case, open-response

[4]Slightly more fourth-grade teachers than eighth-grade mathematics teachers reported positive effects of portfolios and performance events in their schools.

items and portfolios were cited most often, while multiple-choice items were cited by very few teachers (Table 3.1). Performance events, however, were cited by relatively few teachers (only 19 percent) as having a great deal of positive instructional effect—an important finding given that presumed instructional effects are an important motivation for the inclusion of performance events in the assessment.

Only one KIRIS component was cited by more than 10 percent of teachers as having had a great deal of negative instructional impact: Thirty percent of teachers reported that portfolios have had a great deal of negative impact on instruction. In response to open-ended questions about instructional effects, a number of teachers voiced concern about the time involved in writing and compiling portfolios. They also commented on the need to reduce course content because of portfolio time requirements. Teachers' narrative responses to questions about the effects of portfolios are more fully described later in this report.

When asked how useful the *information from* each cognitive component is for improving instruction in their classes, teachers' rankings roughly paralleled their views on positive instructional effects. Across the two grades, the open-response items and portfolios were most often cited as providing very useful information (see Table 3.2).[5] Performance events and multiple-choice items were said to provide very useful information for instructional improvement by fewer teachers.

EDUCATORS' JUDGMENTS ABOUT KIRIS AS A MEASUREMENT AND ACCOUNTABILITY TOOL

The surveys probed educators' views of three aspects of the adequacy of KIRIS as a measurement tool: the adequacy of the test itself (e.g., the domain it assesses and the way it is administered and scored), the accuracy of the information it yields about student performance, and its accuracy as a measure of school effectiveness. The survey also asked for views about rewards and sanctions based on KIRIS.

Table 3.1

Percentage of Teachers Reporting That Each Cognitive Component of KIRIS Has Had "a Great Deal" of Positive or Negative Effect on Instruction in Their Schools

	Perceived Positive Effects	Perceived Negative Effects
Multiple-choice items	6	3
Open-response items	43	6
Performance events	19	5
Portfolios	40	30

NOTE: Fourth-grade teachers were asked about instruction generally, while eighth-grade mathematics teachers were asked about instruction in mathematics.

[5]Note that the percentages in Table 3.2 are not strictly comparable to those in Table 3.1. The questions reflected in Table 3.1 had four response options (not at all, a small amount, a moderate amount, a great deal), while those reflected in Table 3.2 offered three choices (not at all useful, somewhat useful, very useful). Hence, it is reasonable to compare rankings across the two tables but risky to compare the corresponding percentages themselves across the tables.

Table 3.2

Percentage of Teachers Reporting That the Information from Each Cognitive Component of KIRIS Is "Very Useful" for Improving Instruction in Their Classes

Multiple-choice items	16
Open-response items	37
Performance events	21
Portfolios	31

Support for the Testing Domain, Administration, and Scoring

Despite generally positive views of the role of KIRIS in instructional improvement, teachers' opinions about the adequacy of the KIRIS assessment were clearly mixed. Both negative and positive comments about the assessment elicited agreement from majorities of teachers. More expressed strong views of the limitations of KIRIS than of its strengths.

The majority of teachers reported some degree of agreement with several positive statements about KIRIS, but relatively few expressed strong agreement. About three-fourths of the teachers (77 percent) agreed that KIRIS tests a wider range of skills than multiple-choice tests, but only a modest number (27 percent) reported strong agreement. An even more striking example was a statement that KIRIS tasks are based on realistic situations (an important argument of proponents of performance assessment): 70 percent of teachers agreed, but only 8 percent agreed strongly. Over half of the teachers (57 percent) agreed that "KIRIS assessments more closely resemble what I teach than do standardized tests," but fewer than 10 percent strongly agreed.

In contrast, many teachers expressed strong agreement with a number of negative comments about KIRIS. First, about half of the teachers (53 percent) reported strong agreement that scoring standards for KIRIS are inconsistent across years and disciplines. And nearly half (45 percent) strongly agreed that the curriculum content for the assessments is not defined well enough for them to prepare students adequately.

Second, many teachers also reported strong negative views about the role of writing in the assessment. Forty-five percent of teachers strongly agreed that poor writing skills make it hard to judge some students' mathematics achievement, and 32 percent strongly agreed that good writing skills make it appear that some students know more mathematics than they do.

Third, there were two issues that more fourth-grade teachers than eighth-grade mathematics teachers flagged as potential limitations of the assessment. Forty percent of fourth-grade teachers but only 15 percent of eighth-grade mathematics teachers strongly agreed that assessment materials and testing requirements are developmentally inappropriate for students in their grade. Similarly, 31 percent of fourth-grade teachers but only 12 percent of eighth-grade mathematics teachers strongly agreed that testing times for some KIRIS tasks are too short for students to show how well they can perform.

Finally, many teachers expressed concern about the adequacy of guidelines for the portfolio component of KIRIS. Fewer than 5 percent of teachers strongly agreed that the

guidelines are sufficiently standardized to make portfolios reasonable indicators of students' achievement. In fact, about half of the eighth-grade mathematics and a third of the fourth-grade teachers strongly disagreed with this statement.

Perceived Accuracy of Student Achievement Information

Notwithstanding these concerns about attributes of KIRIS that could potentially distort the information KIRIS provides about student performance, the majority of teachers expressed positive opinions about the accuracy of student achievement information provided by KIRIS. Our surveys asked teachers explicitly about their perceptions of the accuracy of student achievement information provided by each of the four cognitive components. Fourth-grade teachers were asked about accuracy of information about students' achievement generally and about the adequacy of student achievement information in mathematics, reading, science, and social studies. Eighth-grade mathematics teachers were asked about achievement in mathematics.

Across the cognitive components, between 52 percent and 81 percent of teachers said the student achievement information KIRIS provides is "somewhat" or "very" accurate (Table 3.3). (Few teachers—less than 11 percent—described the achievement information provided by any of the KIRIS assessments as "very accurate.") The multiple-choice items in the transitional assessment were cited by the most teachers (81 percent) as providing accurate information. The open-response items on the transitional assessments were said to provide accurate information by 62 percent of teachers. For both components, slightly more eighth-grade mathematics teachers than fourth-grade teachers said the information is accurate. About half of the teachers said mathematics portfolios (54 percent) and performance events (52 percent) provide accurate information about students' achievement. About the same percentage of fourth-grade teachers (60 percent) said information from writing portfolios is accurate.

The majority of fourth-grade teachers reported that the information provided by KIRIS in each subject area is adequate. The reading assessment was most often cited as providing accurate information. Sixty-three percent of fourth-grade teachers reported that the information KIRIS provides about students' achievement in reading is somewhat or very

Table 3.3

Percentage of Teachers Reporting That Student Achievement Information from Each KIRIS Cognitive Component Is "Somewhat" or "Very" Accurate

Multiple-choice items	81
Open-response items	62
Performance events	52
Mathematics portfolios	54
Writing portfolios	60

NOTE: Fourth-grade teachers were asked about student achievement generally, while eighth-grade mathematics teachers were asked about achievement in mathematics. Fourth-grade teachers were asked separately about single-subject and interdisciplinary performance events, and eighth-grade mathematics teachers were asked about mathematics performance events. Only fourth-grade teachers were asked about writing portfolios.

accurate; about half agreed the information KIRIS provides in social studies, mathematics, and science is accurate. Most respondents indicated that KIRIS information on students' achievement is somewhat, rather than very, accurate.

Elementary- and middle-school principals reported positive views of the usefulness of student performance information KIRIS provides to them. Three-quarters of principals said KIRIS provides them and other school principals with more than slightly useful (either "somewhat" or "very" useful) information about student performance; a quarter said the information is very useful.

Perceived Accuracy of School Effectiveness Data

KIRIS data are used to draw conclusions about the effectiveness of educational programs. We asked both principals and teachers numerous questions specifically about this use because a test can provide good information about student performance even if it does not provide a valid basis for conclusions about educational effectiveness. For example, comparisons among schools could reflect differences in students' backgrounds rather than differences in educational effectiveness. Conversely, an assessment may provide adequate school-level information about student achievement—not the same as information about school effectiveness—while yielding only inadequate information about the performance of *individual* students because of matrix sampling of test items. The survey asked educators for their opinions of the reasonableness of KIRIS data for drawing inferences about school effectiveness and asked them to compare KIRIS in this respect to traditional standardized, multiple-choice achievement tests. Both teachers and principals also were asked to respond to questions about a number of specific factors that might distort comparisons among schools or measures of change over time. Although the survey did not explicitly point this out to respondents, the factors about which we asked have little to do with the format of KIRIS; most would apply similarly to traditional multiple-choice tests if they were used in the same way.

Echoing their statements about the usefulness of KIRIS information about student achievement, principals and teachers reported generally positive opinions of the value of KIRIS for judging the effectiveness of schools. A large majority of principals (88 percent) said KIRIS results are reasonable for making inferences about school improvement; about a quarter judged them to be very reasonable for this purpose (24 percent). The teachers generally concurred.

In every case, the percentage of teachers who reported that each of the KIRIS cognitive components is reasonable for drawing conclusions about school effectiveness was as high or higher than the percentage reporting that the same component provides accurate information about student performance (Table 3.4; compare with Table 3.3). A majority of teachers said that each of the components is reasonable as an indicator of school effectiveness. (Fewer than 10 percent of teachers said the data from any one of the assessments are "very" reasonable for this purpose.) The large majority of teachers (87 percent) reported that results provided by the multiple-choice items in the transitional

assessment are reasonable for drawing conclusions about the effectiveness of educational programs. Somewhat fewer agreed that the performance-based components are reasonable for this purpose. Slightly more fourth-grade teachers than eighth-grade mathematics teachers reported that portfolios provide reasonable information about educational effectiveness. When asked about the cognitive indicators *taken together*, 72 percent of teachers said they provide somewhat or very reasonable information about the effectiveness of schools' educational programs.

Teachers' judgments of the reasonableness of information provided by the noncognitive indicators parallel those for the cognitive components. Three-quarters judged the noncognitive indicators *taken together* as reasonable indicators of educational effectiveness. Seventy-three percent said attendance rates provide reasonable information. Sixty-two percent of teachers at both levels agreed that promotion rates provide reasonable information, and roughly the same percentage of eighth-grade mathematics teachers responded similarly to dropout rates.

The surveys did not ask principals to offer views on the individual components of KIRIS. Principals, however, were divided in their opinion of the value of KIRIS data compared with data provided by standardized, multiple-choice tests. Forty-four percent of principals somewhat or strongly agreed that KIRIS provides a better view of educational effectiveness than standardized multiple-choice tests (slightly more elementary than middle-school principals expressed this view); the rest disagreed. Similarly, fewer than 40 percent of teachers agreed that KIRIS better reflects educational effectiveness than do standardized, multiple-choice tests.

Although most respondents voiced positive views when asked general questions about the reasonableness of KIRIS as an indicator of school effectiveness, many believe that factors other than educational effectiveness influence score differences among schools. About a third of principals (32 percent) and 60 percent of teachers strongly agreed that score differences between schools often reflect students' characteristics more than school effectiveness. Over 60 percent of both groups strongly agreed that schools with highly transient populations are

Table 3.4

Percentage of Teachers Reporting Information from Each Component of KIRIS Is "Somewhat" or "Very" Reasonable for Drawing Conclusions About Educational Effectiveness

Multiple-choice items	87
Open-response items	73
Performance events	62
Portfolios	62
Cognitive indicators taken together	72
Attendance rates	73
Promotion rates	62
Dropout rates	59
Noncognitive indicators taken together	75

NOTE: Only eighth-grade mathematics teachers were asked about dropout rates.

at an unfair disadvantage on KIRIS. Further, a fifth of the principals and over 35 percent of teachers strongly agreed that score changes mix together school improvement with year-to-year changes in student cohorts.

These educators also questioned the degree to which score differences reflect educationally important performance differences. Almost 40 percent of principals and about half of the teachers (52 percent) strongly agreed that some schools (not necessarily their own) have found ways to raise KIRIS scores without really improving education. (Eighty-seven percent of teachers and 71 percent of principals expressed some degree of agreement with this statement.) Just over a third of principals (37 percent) and about half of the teachers (52 percent) strongly agreed that KIRIS gains during the first biennium were sometimes misleading because some schools aimed for poor performance in the baseline years. (This potential distortion was only an issue during the first biennium.) Only about 15 percent of principals and teachers strongly agreed that comparisons among schools may be distorted because some schools retain students in nonaccountability grades to improve their index.

Finally, teachers suggested that several additional factors may distort KIRIS comparisons. We asked Kentucky teachers (but not principals) whether three additional factors *did not* distort comparisons between schools and over time. (We used this construction to balance positive and negative survey prompts.) A large majority of teachers (85 percent) reported that differences in student motivation may distort KIRIS comparisons across schools or years. (That is, they disagreed with a statement that such differences *do not* distort comparisons.) Eighty-one percent indicated that differences in access to test preparation materials may distort comparisons between schools and over time, and 64 percent suggested that test administration differences may distort comparisons.[6]

Support for Rewards and Sanctions

As earlier noted, KIRIS results are used to hold schools accountable. Results are published, and the state uses the data to identify schools that show improvement for rewards and (in the future) schools that do not show improvement for sanctions. We asked educators about their opinions of the use of KIRIS for accountability.

Few principals and teachers expressed support for rewards and sanctions. Only about a quarter said they somewhat or strongly support the imposition of rewards and sanctions on the basis of KIRIS scores (27 percent). Thirty-four percent of principals and 44 percent of teachers said they strongly oppose rewards and sanctions. Further, 67 percent of teachers strongly agreed (almost all agreed to some degree) that rewards and sanctions will unfairly reward and punish many teachers.

---

[6]Note that these are the percentages of teachers "somewhat" and "strongly" disagreeing with negatively phrased statements about KIRIS. These results are therefore not comparable to the percentages "strongly" agreeing with positively phrased statements reported in the preceding paragraphs.

4. IMPACT OF KIRIS ON SCHOOL MANAGEMENT AND INSTRUCTION

Many educators in Kentucky had strong opinions about the impact of KIRIS on schooling. They reported numerous changes in the content and delivery of education, many of which are consistent with the goals of the reform program. Teachers reported that KIRIS had increased the constructive feedback principals give to the teachers, had given teachers more voice in curricular decisions, and had encouraged teachers to experiment in their teaching and to accommodate students with different learning styles. Educators reported increased expectations for students, more use of open-response test questions, less use of multiple-choice assessments, more group work and writing, more emphasis on problem-solving, and less emphasis on computation and language mechanics. They reported that KIRIS has reduced the use of "pull-out" programs and special classes for remedial work and promoted the use of time outside the normal school day for remedial work (e.g., before and after school and during summers).

However, educators also reported that KIRIS has caused high stress. Most teachers strongly agreed that KIRIS has put teachers under undue pressure. Most teachers reported that teacher morale in their schools is low and has been harmed by KIRIS, and about half reported that KIRIS has reduced their own job satisfaction. A sizable minority reported that KIRIS has also decreased the morale of their students.

THE CHANGING ROLE OF THE PRINCIPAL

Most principals reported that KIRIS had substantially changed their role as the instructional leader of their schools. A large majority of principals (81 percent) reported that the percentage of time they devote to instructional issues had increased as a result of KIRIS (51 percent strongly agreed with this statement), and all principals reported that in response to KIRIS, they had encouraged their teachers to "improve instruction generally."

Most teachers (79 percent) reported that the principal in their school had provided constructive feedback to the teachers on their performance. Moreover, roughly half (45 percent) of the teachers reported that KIRIS had increased this feedback.

Principals reported giving a number of different types of rewards to teachers specifically for their students' good performance on KIRIS. Overall, 70 percent of school principals publicly recognize teachers within the school (Table 4.1). In addition, 57 percent of the principals reported recognizing teachers outside of the school for their students' KIRIS performance. Half reported giving teachers priority for materials, and 67 percent reported giving teachers additional resources for use within the school as a reward for their students' KIRIS performance. About one-fourth of middle-school principals said they gave teachers greater choice in the classes they teach as a reward for the good performance of students.[7]

[7]This question was not asked of elementary principals.

Table 4.1

Percentage of Principals Reporting Rewarding Teachers Based on Their Students' KIRIS Scores

Type of Reward or Recognition	Fourth Grade	Eighth-Grade Mathematics	Combined
Public recognition within the school	67	73	70
Additional resources for use within the school	71	63	67
Public recognition outside the school	53	61	57
Priority on requests for materials	47	53	50
Greater choice of classes to teach	—	27	—
Relief from administrative or disciplinary duties	12	22	17
Greater choice of students to teach	19	13	16
A lighter teaching load	16	14	15
Extra pay	10	8	9

Only a few principals reported using other rewards about which we asked, including extra pay, a lighter teaching load, and relief from administrative and disciplinary duties.

KIRIS also appears to have influenced staffing decisions made by a sizable minority of principals. Overall, 37 percent of elementary-school principals and half of middle-school principals reported moving teachers among grades because of KIRIS (either into a tested grade, out of a tested grade, or both). Approximately one-fourth of the elementary principals (27 percent) reported transferring good teachers into fourth-grade classes, and one-fourth reported transferring weaker teachers out of fourth-grade classes. The corresponding numbers for middle-school principals and eighth-grade mathematics teachers were larger—36 percent and 38 percent, respectively.

Principals also reported that KIRIS had affected attrition. We asked separately about both good and poor teachers leaving teaching because of KIRIS. The percentage of principals who reported that more good teachers had left teaching than previously because of KIRIS (23 percent) was strikingly similar to the percentage reporting that more mediocre or poor teachers had left because of KIRIS (20 percent).

EXPECTATIONS FOR STUDENT ACHIEVEMENT

Principals reported taking the "high-standards-for-all" message seriously. We asked principals to indicate the degree to which they encouraged their teachers to focus on setting higher expectations for various groups of students in response to KIRIS using a Likert response scale with the options "not at all," "somewhat," and "a great deal." Most reported encouraging their teachers "a great deal" to focus on setting higher expectations for students. Furthermore, there was little difference in the percentage of principals reporting "a great deal" for low achieving (78 percent), average (84 percent), and high-achieving students (75 percent).

Slightly fewer teachers (68 percent) reported that expectations had changed since KIRIS began. Of those teachers who reported a change in expectations, most reported an increase in expectations for all of the groups about which we asked. More than 80 percent of these teachers reported that expectations had increased "somewhat" or "a great deal" for low-, average-, and high-achieving students, and 74 percent said the same of special-education students. Since only 68 percent of teachers reported a change in expectations, however, the percentage of *all* teachers who reported an increase in expectations ranged from 50 percent (for special-education students) to 60 percent (for average students).

A less equitable picture emerges when one looks at the percentage of teachers reporting that expectations had increased a great deal. Few teachers reported a great deal of increase in expectations for any group. However, more teachers felt expectations had increased greatly for high-achieving students than for average students, low achievers, and special-education students (see Table 4.2). Except for special-education students, these differences are small, but they are consistent across grades.[8]

Evidence of more positive effects for higher-achieving students also appeared in a set of questions that asked teachers about changes in the emphasis on high standards in their schools. Sixty-two percent of teachers reported that the emphasis on high standards had changed for some students in their schools. Of the teachers that reported such a change, far more considered the change helpful than harmful for each of the four groups about which we asked: special-education, low-achieving, average, and high-achieving students (Table 4.3). However, considerably more teachers (particularly those teaching fourth-grade classes) thought the emphasis on high standards has been helpful for average students and high-achieving students than for special-education and low-achieving students. In addition, a small but appreciable minority of the fourth-grade teachers reported that the emphasis on high standards has been somewhat or very harmful for special-education and low-achieving students.

Table 4.2

Percentage of All Teachers Reporting That Academic Expectations Have "Increased Greatly," by Grade and Type of Student

Type of Student	Fourth Grade	Eighth-Grade Mathematics	Combined
Special education	14	10	12
Low achieving	18	14	16
Average achieving	20	17	18
High achieving	24	24	24

[8]In addition, a similar pattern was found in Maryland teachers' responses to questions about the Maryland School Performance Assessment Program (Koretz et al., 1996).

Table 4.3

Percentage of Those Teachers Reporting a Change in Emphasis on High Standards Who Deemed That Change Harmful or Helpful

Type of Student	Fourth Grade		Eighth-Grade Mathematics		Combined	
	Harmful	Helpful	Harmful	Helpful	Harmful	Helpful
Special education	20	44	13	62	17	53
Low achieving	27	56	6	67	17	61
Average achieving	10	80	8	83	9	81
High achieving	6	86	11	73	8	80

COURSE OFFERINGS

We asked principals a number of questions about course offerings, and an additional caveat about these specific responses should be noted. Survey questions are not fully adequate to characterize changes in student coursework; changes that appear dramatic might be only superficial in reality, and visa versa. For instance, a school might decide to eliminate a course called remedial math and place students who previously would have been assigned to remedial math in the general math class. This could be a major shift for the school that raises the level of content delivered to the lowest achieving students, or it could be no more than a name change, with few implications for the students involved. Thus, the survey results reported here are only a first step in getting a picture of the changes in course offerings stemming from KIRIS and the broader school reform effort.

About half of the middle school principals (48 percent) report that KIRIS has affected course offerings in their schools. When asked for the names of specific courses that had been added or dropped, many principals indicated that they had added courses without dropping any. Courses in mathematics (e.g., algebra, pre-algebra, geometry, extended math, higher-level mathematics) were the most common additions (55 percent of principals that noted specific additions included at least one math class in their list of additions). Almost as popular were courses in writing, writing portfolios, or journalism, with 45 percent reporting additions. A number of principals also noted adding courses in social studies (or Kentucky history—18 percent), computers (or keyboarding—18 percent), or arts and humanities (18 percent).

Principals generally reported dropping courses in either of two areas: basic level or remedial courses in core academic subjects (e.g., remedial math, remedial English, basic social studies) and enrichment classes such as music, foreign languages, and physical education (each noted by only a couple of principals). (As noted below, the provision of remedial services outside of regular class time reportedly increased in many schools.) Sixty percent of the principals who listed specific courses that had been dropped included at least one basic level course in a core subject; 33 percent listed a basic math course that had been dropped.

Roughly half of the principals (51 percent) reported that KIRIS has affected the courses offered in the high schools in their area. However, it is likely that many middle-school principals have only limited information about course offerings in the high schools in

their area. (In fact, when prompted for specific course additions or deletions that have occurred in area high schools, many principals responded with "I don't know.") Of those who gave specifics additions, principals most commonly noted higher-level math courses. The course deletion at the high school level noted by the most principals was vocational education.

STUDENT GROUPING AND REMEDIATION

Kentucky educators reported a trend away from homogeneous grouping of students but an increase in the time students spend on remedial work—mostly outside the normal school day. The changes differ somewhat by grade level.

Elementary principals reported that homogeneous grouping for academic classes was rare and had decreased markedly since KIRIS began. Very few elementary principals (12 percent) reported grouping students homogeneously for their academic classes. Of the great majority of principals who reported no current homogeneous grouping, almost half (46 percent) reported having had such grouping before the KIRIS system was put in place. Fourth-grade teachers also noted a decline in homogeneous grouping within classes. Approximately three-fourth (72 percent) reported having students work in groups of mixed ability more often than before KIRIS was first administered.

However, a substantial number of elementary principals reported changes that suggest more differential assignment of materials based on the achievement level of students. Forty-five percent indicated that more students are given advanced materials in response to KIRIS, and an almost equal number (41 percent) reported that more students are given remedial services in response to KIRIS. The most dramatic reported change in remedial services pertained to the number of student in before- or after-school programs. Eighty-two percent of elementary principals reported an increase in the number of students in before- or after-school remedial programs as a result of KIRIS, and 68 percent reported an increase in the number of students in summer remedial programs. Thirty-eight percent of principals reported that test scores (on the continuous assessment, scrimmage tests, or KIRIS itself) are used in assigning students to remedial programs.

Homogeneous grouping is more common at the middle-school level: Half of the middle school principals reported grouping students homogeneously for one or more of their academic classes. Nonetheless, there are indications that grouping has decreased as a result of KIRIS. Forty-one percent of the principals who reported using homogeneous grouping indicated that they use it less in response to KIRIS. Moreover, of the principals who did not report current homogeneous grouping, about one-fifth reported using it before the KIRIS implementation.

Additionally, 22 percent of eighth-grade principals reported that their schools had increased the number of students assigned to advanced classes as a result of KIRIS, while very few (6 percent) indicated that their schools had decreased these assignments.

Many middle-school principals, like many elementary principals, reported an increase in participation in before- or after-school remedial programs (78 percent) and in summer remedial programs (53 percent). Forty-five percent of principals reported using KIRIS scores

(on the assessment itself, the scrimmage tests, or the continuous assessments) to assign students to remedial services.

Eighth-grade mathematics teachers also indicated that their use of homogeneous grouping had declined. Thirty-eight percent reported that they grouped students homogeneously based on ability within their classes less than they did before KIRIS. Also, half reported having students work in groups of mixed ability more often than before KIRIS was first administered.

EFFECTS ON CLASSROOM ASSESSMENTS

Teachers reported making a number of changes in the classroom assessments they administered to their students since KIRIS was implemented. Almost 50 percent reported decreasing their use of multiple-choice assessments. The majority of teachers reported increased use of virtually all other types of assessment tasks about which we asked (Table 4.4). Fifty-eight percent of fourth-grade teachers attributed these changes primarily to KIRIS, and an additional 37 percent attributed them to both KIRIS and the widely discussed reforms spurred by the National Council of Teachers of Mathematics (NCTM). Among eighth-grade teachers, 39 percent attributed the changes to KIRIS, and 51 percent attributed them to both KIRIS and NCTM.

EFFECTS ON INSTRUCTION

Has KIRIS influenced the way that teachers teach? And, if so, what are the most prevalent changes teachers have made? These were two of the key research questions that motivated this study. We asked teachers a number of questions pertaining to changes in practice: both closed-ended questions that would indicate the percentage of educators perceiving specific instructional changes and open-ended questions aimed at getting educators' opinions of the most salient changes.

Teachers reported that their efforts to improve instruction and learning have increased. Virtually all teachers (93 percent) reported that they have focused at least a moderate amount on "improving instruction generally" in their efforts to improve scores on KIRIS. Also, 87 percent of teachers reported that they were encouraged to experiment in their teaching. Forty-three percent responded that KIRIS had led to an increase in the degree to which they are encouraged to experiment.

Changes in Curricular Emphasis

Ninety percent of fourth-grade teachers and 87 percent of eighth-grade mathematics teachers reported focusing "a moderate amount" or "a great deal" on improving the match between the content of their instruction and the content of KIRIS. Sixty-nine percent of teachers indicated that there was content they emphasized more because of KIRIS. Principals appeared to be facilitating a change in the curriculum taught in their schools. Seventy percent of the principals reported encouraging their teachers "a great deal" to focus

Table 4.4

Percentage of Teachers Reporting Increased Use of Various Classroom Assessment Types by Grade

Fourth Grade	
Open-response tasks requiring:	
a numerical answer	35
a few words up to a sentence	67
more than a sentence	83
a table, chart, etc.	72
Open-response tasks requiring:	
less than 5 minutes	45
5–30 minutes	82
more than 30 minutes	75
Group tasks	
group work yielding an individual product	84
group work yielding a group product	85
Eighth Grade	
Open-response tasks requiring:	
a numerical answer	34
a few words or sentences	74
a paragraph or more	91
a table, chart, etc.	67
Open-response tasks requiring:	
less than 5 minutes	40
more than 5 minutes but less than a class period	76
an entire class period	74
more than one class period	67
Group tasks:	
group work yielding an individual product	64
group work yielding a group product	70

instruction on "skills or content likely to be on KIRIS." Most (93 percent) reported that their schools' emphasis on material likely to be emphasized by KIRIS had increased; 46 percent said it had increased greatly.

The other side of alignment is a perceived decrease in emphasis on material not emphasized by the assessment. Only a minority of principals reported deemphasis on untested materials, but most teachers did. Principals reported that there had been a decrease in their schools in the emphasis given to each of the following areas: the pre-KIRIS curriculum (33 percent), untested subject areas (44 percent), and material unlikely to be tested even though it is in a tested subject area (32 percent). In contrast, 88 percent of fourth-grade teachers agreed that "KIRIS has caused some teachers to deemphasize or neglect untested subject areas," and 40 percent strongly agreed with this statement. Eighth-grade mathematics teachers were asked a narrower question: Had KIRIS caused some mathematics teachers to deemphasize or neglect untested mathematics topics? The responses, however, were similar: 86 percent agreed, and 43 percent strongly agreed, that such changes had occurred. Fifty-four percent indicated that there is content they themselves emphasized less because of KIRIS.

In addition, approximately three-fourths of the teachers (77 percent) responded that too much of their time was diverted from instruction to deal with classroom management issues. In addition, roughly half (49 percent) responded that KIRIS had led to an increase in the amount of time that was diverted.

We found a greater agreement concerning what specific content and activities had been increased than we did about what had been decreased. This is not surprising, given that KIRIS clearly emphasizes certain skills and content areas and that teachers reported increasing the instructional time they devoted to those areas. However, what teachers take time away from to accommodate KIRIS is not clearly dictated by the test. Therefore, teachers reported taking time away from a wide variety of content areas as well as various classroom activities.

Because elementary and middle schools are organized differently, teachers and principals at the two levels have different opportunities to change the time allocated to various content and activities. Accordingly, we asked different questions by grade concerning specific changes in emphasis, and the results are reported separately by grade.

Fourth-Grade Teachers

Fourth-grade teachers were asked how they have shifted available instructional time *between* subject areas since KIRIS was first administered. The subject area to which by far the most teachers indicated allocating more instructional time was writing (Table 4.5). Virtually all teachers reported allocating more time to writing, and 83 percent indicated the amount of time allocated to writing had "increased greatly." For every subject area about which we asked except writing, a substantial minority of the teachers indicated that instructional time had decreased. The subject areas for which the most teachers indicated a decrease since KIRIS began were art, social studies, science, and reading. Eighty-nine percent of the teachers indicated that these changes were due largely to KIRIS.

Fourth-grade teachers were also asked how they have changed emphasis *within* mathematics and language arts (Table 4.6). Within language arts, most fourth-grade teachers reported an increase in emphasis on writing for a variety of purposes. A decrease in emphasis was reported on spelling, punctuation, and grammar. Most fourth-grade teachers indicated that they have increased the emphasis on mathematics communication and

Table 4.5

Percentage of Fourth-Grade Teachers Reporting Changes in Content Emphasis

Content Area	Decreased	Increased
Writing	0	95
Mathematics	13	45
Reading	28	31
Science	30	30
Social studies	33	30
Art	34	10
Music	21	5
Physical education	22	3

Table 4.6

Percentage of Fourth-Grade Teachers Reporting Changes in Content Emphasis Within Language Arts and Mathematics

Content	Decreased	Increased
Language arts		
Writing for a variety of purposes	0	97
Analysis and evaluation of text	13	47
Literary comprehension	14	39
Spelling, punctuation, and grammar	43	17
Mathematics		
Problem-solving using meaningful tasks	1	87
Mathematical communication	2	84
Application	2	65
Number facts and computation	42	4

meaningful problem-solving. The only aspect of mathematics in which a sizable proportion of fourth-grade teachers indicated a decrease in emphasis was number facts and computation.

Because much of the time spent in elementary grades is on education that is not traditional desk work, we also asked about a variety of other activities. A sizable proportion of the teachers reported that their school had reduced the amount of time spent on each of the following activities to accommodate KIRIS: recess (50 percent), organized play (43 percent), student performances (43 percent), student choice time (e.g., games and computer work—43 percent), and class trips (39 percent).

Eighth-Grade Mathematics Teachers

For the most part, eighth-grade mathematics teachers could change emphases only *within* the one subject, and we therefore asked them about only within-subject time allocations.

Most eighth-grade mathematics teachers indicated that they had increased the emphasis on communicating mathematics ideas and solutions and on problem-solving and reasoning (Table 4.7). The only aspects of mathematics for which a sizable proportion of eighth-grade mathematics teachers indicated a decrease in emphasis were computation and algorithms.

Perceived Positive and Negative Instructional Effects

Teachers reported that KIRIS had produced both positive and negative instructional effects. Overall, slightly more teachers reported that there were positive effects than negative effects. Most fourth-grade teachers (90 percent, including even those opposed to KIRIS) reported that at least one of the current parts of the KIRIS assessments (open response, performance events, or portfolios) had had a moderate or great deal of positive effect on instruction, and most eighth-grade mathematics teachers (85 percent) concurred. However, teachers also indicated that KIRIS had had negative effects on instruction: Sixty-nine percent of fourth-grade teachers and 64 percent of eighth-grade mathematics teachers

Table 4.7

Percentage of Eighth-Grade Mathematics Teachers Reporting Changes in Content Emphasis

Content	Decreased	Increased
Communicating mathematical ideas and solutions	2	93
Problem-solving and reasoning	3	82
Use of graphs and tables	2	78
Data analysis	3	72
Space, dimensionality, and measurement	4	63
Conceptual knowledge	15	40
Ratios, proportions, and percentages	16	22
Computation and algorithms	58	5

reported that at least one of the current parts of the assessment had had a moderate or great deal of negative effect on instruction.

Teachers were also asked open-ended questions about the most positive and negative effects of KIRIS on instruction in their schools. The responses to these questions should be interpreted differently from the responses to the closed-ended questions that composed the bulk of the survey. When questions are open-ended, the likelihood is that far smaller percentages of respondents will happen to offer any particular comment. The small percentages of respondents offering a given comment in response to an open-ended question may not indicate that the view expressed is unimportant or held by few people. For example, when asked an open-ended question about the most important negative effects of KIRIS during the phone interview, 14 percent of fourth-grade teachers noted doubts about the developmental appropriateness of KIRIS for their students. In the mail survey, these teachers were also asked to express the strength of their agreement or disagreement with the statement that "assessment materials and testing requirements are developmentally inappropriate for students in my grade." Eighty percent of fourth-grade teachers expressed some degree of agreement with that statement, and 40 percent expressed strong agreement. Thus, although most fourth-grade teachers question the developmental appropriateness of KIRIS for their students, few selected that as an issue upon which to focus when asked to comment on the program's most negative aspects. Accordingly, in what follows, we report percentages as low as 5 percent, even though that corresponds to only five teachers per grade.

The positive comments made by the most fourth-grade teachers and eighth-grade mathematics teachers concerned writing (55 percent and 43 percent of teachers, respectively). In particular, many teachers commented that their emphasis on writing had increased. Teachers also commented that students' writing and communication skills had improved.

A sizable number also commented that KIRIS had led to more focus on problem-solving and thinking skills. Eighteen percent of fourth-grade teachers commented that emphasis on thinking skills had increased or that students' thinking skills had improved. Eight percent made similar comments about problem-solving skills. At the eighth-grade

level, 23 percent of mathematics teachers noted increased emphasis on, or increased student achievement related to, thinking skills, and 14 percent noted increases involving problem-solving skills. Some teachers also commented that KIRIS had led teachers to focus more on real-life applications (6 percent and 24 percent in fourth and eighth grade, respectively), hands-on activities (4 percent and 11 percent in fourth and eighth grade, respectively), and cooperative learning (10 percent and 3 percent in fourth and eighth grade, respectively).

Teachers also noted a number of negative effects that KIRIS had had on instruction. The negative comments most often noted differed somewhat between grades.

No single negative comment was offered by a large percentage of fourth-grade teachers. The negative comments made by the most fourth-grade teachers concerned the amount of stress and pressure on teachers and students because of KIRIS—13 percent reported too much stress on everyone, an additional 6 percent specifically noted too much stress on teachers, and a few commented specifically about too much pressure on students. Thirteen percent of fourth-grade teachers voiced concerned about the time taken away from instruction to prepare for or administer KIRIS. An additional 9 percent noted that content had been reduced because of the time involved in the portfolio component. Eleven percent specifically noted content related to basic skills being decreased because of KIRIS.

Ironically, the amount of writing, about which many teachers expressed positive opinions, was also the subject of negative comments from a much smaller number of teachers. Twelve percent of fourth-grade teachers expressed negative comments about the amount of time students spend writing. Eight percent commented that students are tired of writing or are "burnt out" by all of the writing. When asked specifically about the impact of writing, about half of the fourth-grade teachers (52 percent) strongly agreed that the heavy emphasis on writing in KIRIS has caused some students to become tired of writing.

The negative comments made by the most eighth-grade mathematics teachers concerned time involved in writing and compiling student portfolios. Twenty-three percent of teachers noted that they had reduced the mathematics content covered in their classes because of the time required to do portfolios. Eight percent reported that content had been reduced because of time required for writing. Five percent noted that the time required to do portfolios was out of proportion with the weight of the portfolios in the KIRIS accountability index. Four percent noted that portfolios and writing took too much class time. Eleven percent of eighth-grade mathematics teachers also commented that students had become "burnt-out" because of all of the writing.

A number of eighth-grade mathematics teachers also made comments about the time required for KIRIS but without referring specifically to portfolios. Eleven percent commented that they had reduced the mathematics content covered in their classes because of the time required to prepare for or administer KIRIS. Six percent of teachers noted that kids did not perform their best because the assessment takes too much time and effort.

5. PORTFOLIO PRACTICES

Although portfolios make a relatively small contribution to the accountability index,[9] they are the focus of considerable attention because of the performance-based orientation of KIRIS and because of their prominence in the instructional program. As noted in the procedure section, we surveyed eighth-grade mathematics teachers about mathematics portfolios and fourth-grade teachers about writing portfolios. We report results for the two grades and subjects combined and weight them equally when similarities between the practices of the teachers make it appropriate to do so; otherwise, information about the two grades is reported separately.[10]

Portfolio practices varied widely among fourth-grade teachers and among eighth-grade mathematics teachers, which makes it difficult to interpret comparisons of portfolio scores across classes. Despite these differences, there were strong similarities between the grades in terms of typical (median) portfolio practices. In addition, both groups of teachers reported that portfolios required a substantial amount of preparation time, and that scoring was the most time-consuming aspect of preparation. Teachers at both grade levels also reported changes in curriculum and instruction as a result of the portfolios, including becoming more innovative in planning instructional activities.

TRAINING

Almost all fourth-grade teachers and eighth-grade mathematics teachers participated in training to prepare them for the portfolio component of KIRIS. All eighth-grade mathematics teachers received training on the mathematics portfolios. About 85 percent of fourth-grade teachers received training on the mathematics portfolios and about 95 percent on the writing portfolios.

---

[9]The writing component of the accountability index was based entirely on portfolios in grades 4, 8, and 12 in biennium I. The same is expected to be true in biennium II, when writing will account for 14 percent of the accountability index. The mathematics portfolio program was implemented by the end of biennium I, but portfolio scores were not counted in the accountability index for that period. Mathematics portfolios will compose 30 percent of the mathematics score in grades 8 and 12 in biennium II, and this represents 4.2 percent of the overall accountability index.

[10]There are strong similarities between the goals of the two portfolio programs. Four of the five goals of the writing and mathematics portfolio programs are the same—to promote students' skills, knowledge, and confidence in each subject; to document performance; to integrate instruction and assessment; and to provide a basis for curriculum development. Consequently, it is appropriate in many cases to combine comments about fourth-grade writing portfolios and eighth-grade mathematics portfolios. There also are important differences between the two programs. The writing portfolio program has an additional goal of enhancing students' abilities to communicate to different audiences for different purposes, while the mathematics portfolios are designed to help students gain mathematical power over a variety of concepts and principles. This difference as well as differences in the skill level of students, homework policies, organization of the classroom, and flexibility in curriculum make some direct comparisons between the programs less meaningful.

BEST COPY AVAILABLE

CLASSROOM PRACTICES

There was considerable variation in the amount of classroom time devoted to portfolios by fourth-grade teachers and eighth-grade mathematics teachers. Fourth-grade teachers reported spending between 3 and 40 class hours on writing portfolios in a typical month,[11] and the largest clusters of teachers (of roughly equal size and each representing about 12 to 15 percent of the total) were found at 10, 20, 30, and 40 class hours in a typical month. The range of reported class time devoted to eighth-grade mathematics portfolios in a typical month was similar, but the bulk of teachers were clustered at about 5 hours and about 10 hours per month.

Despite the wide variation within each grade and the differences in classroom organization and subject matter between the grades, the median responses regarding classroom time were similar, with eighth-grade mathematics teachers devoting considerably less time (in absolute terms) but slightly more time (in relative terms) to portfolios than fourth-grade teachers. In a typical month, the median fourth-grade teacher reported spending 20 hours of class time on portfolio-related activities (about an hour per day), which represented about 20 percent of the total instructional time. The median eighth-grade mathematics teacher spent six hours on portfolios in a typical month, which represented a somewhat greater proportion of time (about 30 percent of the available class time). The amount of time devoted to portfolios was not constant throughout the year; in the lightest month, the median teachers in both grades devoted about one-half as much time to portfolios as they did in a typical month; in the heaviest month they devoted about twice as much time.

The typical fourth-grade teacher and the typical eighth-grade mathematics teacher divided their classroom portfolio time in similar ways. About three-quarters of the time was divided roughly equally among three activities: preparing students by teaching specific skills, completing portfolio entries for the first time, and revising or rewriting pieces. About 10 percent of the time was spent organizing and managing portfolios, and the remainder was devoted to preparing students by teaching similar tasks.

Fourth-grade teachers and eighth-grade mathematics teachers reported that revision was required or encouraged for most portfolio entries, and in both grades more revisions occurred on assessment portfolio entries than on working portfolio entries.[12] For example, about 50 percent of fourth- and eighth-grade teachers required students to revise pieces in their working portfolios, and about 40 percent more encouraged students to revise. (The remaining 10 percent either permitted but did not encourage revision or had no policy.) By comparison, 60 percent of the teachers in the two groups required revision of assessment portfolio entries, and 30 percent only encouraged it.

The result of these policies was that most of the entries in both students' working and assessment portfolios were revised one or more times, and assessment portfolio entries were revised slightly more often than working portfolio entries. Table 5.1 shows that over 90

---

[11]There were a handful of responses outside this range.

[12]Students compiled "working" portfolios throughout the year, and prior to the deadline for submission, they assembled final "assessment" portfolios.

Table 5.1

Percentage of Portfolio Entries Revised (Standard Deviation)

Number of Times Revised	Fourth-Grade Writing Portfolios		Eighth-Grade Math Portfolios	
	Percentage of Working Portfolio Entries (SD)	Percentage of Assessment Portfolio Entries (SD)	Percentage of Working Portfolio Entries (SD)	Percentage of Assessment Portfolio Entries (SD)
Not at all	17 (23)	6 (15)	28 (32)	21 (31)
Once	36 (25)	31 (29)	38 (30)	43 (34)
Two or three times	32 (23)	46 (30)	25 (27)	28 (31)
Four or more times	14 (22)	17 (25)	10 (21)	8 (19)

percent of the entries in fourth-grade writing assessment portfolios were revised at least once, and the same was true for about 80 percent of the entries in eighth-grade mathematics assessment portfolios. Eighth-grade students revised their mathematics working portfolio entries less frequently than fourth-grade students revised their writing portfolio entries; about one-quarter were never revised at all. Differences in revision policies between grades may reflect differences in the quality of students' initial work as well as differences in the nature of the writing and mathematics tasks students were performing. The high degree of revision in assessment portfolio entries in both grades is consistent with Kentucky's purposes for portfolios. Assessment portfolio entries are supposed to reflect students' best work rather than their ability to produce work "on demand."

More important, the level of revision varied among classrooms in both grade levels. The standard deviations in Table 5.1 are large in proportion to the mean values, indicating large variations in revision practices within grade level. A concrete example will illustrate the magnitude of the differences among teachers. On average, fourth-grade teachers reported that 63 percent of writing assessment portfolio pieces were revised two or more times, suggesting that multiple revision was the norm. However, about one-third of the fourth-grade teachers reported that one-half or more of the pieces in students' writing assessment portfolios were revised only once, i.e., for students in these classes most pieces were revised just a single time. Moreover, if we compute average revision statistics based on the responses of all teachers in a school and compare these school aggregate responses, the differences remain.[13] These school-level differences in portfolio practices are important in interpreting comparisons among schools, an issue that is discussed in a later section.

One of the questions that is often asked about portfolio entries is "whose work is it?" In this case, it appears that the work represents the students' efforts, almost always supported by teachers and frequently by peers. About 90 percent of fourth-grade teachers and eighth-grade mathematics teachers reported that students frequently or always received

[13]We analyzed school average responses in two subsamples: (a) all schools in which we received two or more completed teacher surveys and (b) all schools in which we received completed surveys from more than one-half of the eligible teachers. Both analyses yielded the same results.

help from teachers in completing or revising entries in their working and assessment portfolios. Fifty-five percent indicated that students frequently or always received help from other students. The incidence of assistance from other adults, either in school or at home, was much lower. Only about 30 percent of teachers said this occurred frequently or always.

The vast majority of the fourth-grade teachers and eighth-grade mathematics teachers varied their assistance based on the needs of the students, and their help was equally likely to be based on writing proficiency, mathematics proficiency, reading level, motivation level, or the student's ability to do the specific task. About 80 percent frequently helped students read and understand materials. More than 60 percent of the fourth-grade teachers frequently helped students express ideas clearly and assisted them with the mechanics of writing. Thirty-five to forty-five percent of eighth-grade mathematics teachers frequently provided the same kind of assistance, and about 40 percent of eighth-grade mathematics teachers also frequently helped students with computation or algorithms. In addition, about one-third of the teachers in both grades indicated they always reminded students of the features of responses that score well, and over 80 percent reminded students of these features at least frequently. This means that each student produced his or her portfolio entries under different conditions, and that portfolio scores reflect this customized environment.

There was substantial variation in the difficulty of tasks contained in student portfolios, with far greater differences in eighth grade than in fourth grade. Three-quarters of the fourth-grade teachers reported that the average difficulty of portfolio entries was about equal for all their students, and one-quarter indicated that high-achieving students' portfolios contained more difficult tasks. For the eighth grade, 43 percent thought that high-achieving students had more difficult tasks in their portfolios. The results were similar when teachers' responses were averaged and tabulations were done at the school level.

The total amount of time devoted to the typical portfolio entry also varied considerably across teachers and, when averaged at the school level, across schools. In fourth grade, on average, students spent more than an hour completing work on 58 percent of the tasks in the writing assessment portfolios. However, there were considerable differences in reported completion time among fourth-grade teachers and among eighth-grade mathematics teachers. For example, for fourth-grade assessment portfolios, the two largest clusters of classrooms were at the opposite ends of the time-to-complete scale: In 13 percent of the classes students *never* devoted more than an hour to an assessment portfolio entry and in 25 percent of the classes they devoted more than an hour to *every* assessment portfolio entry. As one would expect, students typically devoted slightly more time to each assessment portfolio entry than to each working portfolio entry, and eighth-grade mathematics students devoted considerably more time than fourth-grade students to entries in their working and assessment portfolios. In eighth grade, on average, only 29 percent of the tasks in the mathematics assessment portfolios were completed in one hour or less.

Fourth-grade teachers and eighth-grade mathematics teachers reported that students worked on their working portfolio entries both in class and outside of class, and fourth-grade students were more likely than eighth-grade students to complete the work wholly or

primarily in class. Forty-one percent of fourth-grade teachers required students to do all the work in class, and an additional 33 percent required students to complete the work primarily in class. The comparable figures for eighth grade were only 10 percent and 19 percent, respectively. Much of the work in eighth-grade students' portfolios was completed outside of class.

Selection of assessment portfolio pieces was primarily the responsibility of students with some input from teachers, although there was considerable variation among fourth-grade teachers and among eighth-grade mathematics teachers in practices for assembling assessment portfolios. On the one hand, most teachers did not give students complete discretion to select all the entries themselves; the typical (median) response was that 20 percent of the entries in fourth-grade writing assessment portfolios and 10 percent of entries in eighth-grade mathematics assessment portfolios were selected by students alone. On the other hand, one-third of fourth-grade teachers and 40 percent of eighth-grade mathematics teachers said students selected one-half or more of the portfolio entries on their own. More common were mixed situations in which students made the choice with input from teachers. Only rarely did teachers take sole responsibility for selection; 5 percent of fourth-grade teachers and 14 percent of eighth-grade mathematics teachers reported they had sole responsibility for selecting a fraction of the pieces (10 percent or more) in the assessment portfolios.

Despite the differences between the subjects and grades, the criteria that were used to select pieces for the fourth-grade writing and eighth-grade mathematics assessment portfolios were quite similar, and they were strongly related to the scoring guides. As Table 5.2 shows, the most important characteristics were that pieces demonstrate achievement on all the scoring criteria and be neat and polished in appearance. Interestingly, fourth-grade teachers and eighth-grade mathematics teachers reported that it was less important to select pieces that might score well on only one or two dimensions. One-quarter of the teachers said this was very important, compared with 40 percent to 50 percent who said it was very important that entries score well on all dimensions. Since portfolios receive a single overall score, this may be a reasonable selection criterion. For the fourth grade (and to a somewhat lesser degree for eighth-grade mathematics), teachers also assigned great importance to selecting entries that showed growth in performance. For the eighth grade, mathematics teachers also gave great importance to entries that demonstrated curricular breadth. Considerably less emphasis was given to selecting tasks that were difficult or challenging, novel or different, involved group work, or were similar to examples from training.

Almost all of the work in students working and assessment portfolios was completed during the school year rather than in previous grades, although there were minor inconsistencies in teachers' reports about this topic. On average, teachers said fewer than 10 percent of working or assessment portfolio entries had been started (or completed) in an earlier year. This was true despite the fact that nearly all principals reported that students' portfolios stayed with them from year to year and that they were encouraging teachers in nonaccountability grades to use portfolios in mathematics and writing.

Table 5.2

Important Criteria for Selecting Assessment Portfolio Entries

Feature	Percentage of Fourth-Grade Teachers Indicating "Very Important" (Writing Portfolios)	Percentage of Eighth-Grade Math Teachers Indicating "Very Important" (Math Portfolios)
Show curricular breadth	NA	54
Show growth in performance	57	43
Show achievement/scorable on all dimensions	48	39
Neat and polished in appearance	43	34
Exemplary/score high in one or two areas	22	27
Novel or different	14	13
Similar to best examples from training	13	13
Include group project work	5	7
Difficult or challenging	4	23

The mathematics assessment portfolios contained primarily written work, but they occasionally included a variety of other materials as well.[14] One-third of the eighth-grade mathematics teachers said that more than 10 percent of the assessment portfolios in their classes contained models or constructions. About one-quarter indicated that 10 percent or more assessment portfolios contained pictures or photographs, and about one-quarter reported that 10 percent or more contained other nonwritten work. All of these instances were rare; fewer than 12 percent of the eighth-grade mathematics teachers reported finding each type of nonwritten materials in more than one-half of the mathematics portfolios. Computer programs appeared in at least 10 percent of the assessment portfolios in about 20 percent of the classes, and videos were almost entirely absent.

Portfolio entries are factored into final report card grades by most fourth-grade teachers and most eighth-grade mathematics teachers. About 90 percent of eighth-grade mathematics teachers included portfolio entries in the computation of final mathematics grades, while only about 60 percent of the fourth-grade teachers counted portfolio entries when assigning final grades in language arts. Among teachers who considered portfolio entries when assigning grades, about three times as many considered both working and assessment portfolio entries as considered only working portfolio entries. Almost no teacher counted just assessment portfolio entries in report card grades. Table 5.3 shows the distribution of credit given to portfolio entries in student grades.

[14]This question was not included in the fourth-grade survey.

Table 5.3

Credit for Portfolio Work in Student Grades

Amount of Credit Given Portfolio Work in Student Grades	Percentage of Fourth-Grade Teachers (Writing Portfolios)	Percentage of Eighth-Grade Math Teachers (Math Portfolios)
Not counted in the final grade	38	9
1% to 10% of the final grade	14	26
11% to 25% of the final grade	21	40
26% to 50% of the final grade	12	13
51% to 75% of the final grade	8	7
More than 76% of the final grade	0	0
Different credit for different students	8	5

BURDENS

The portfolio assessment placed additional demands on fourth-grade teachers and eighth-grade mathematics teachers to prepare lessons and materials, identify appropriate tasks, and score students' assessment portfolios, and both groups of teachers reported that the burdens had not declined from previous years. Teachers spent considerable time outside of class preparing for the portfolios. The median fourth-grade teacher and the median eighth-grade mathematics teacher devoted 10 hours to preparation for the portfolios in a typical month, and preparation time ranged from 5 hours in the lightest month to 23 hours in the heaviest month. Table 5.4 shows that fourth-grade teachers and eighth-grade mathematics teachers reported spending their preparation time in generally similar ways. For example, eighth-grade mathematics teachers spent almost one-half of their out-of-class time scoring or evaluating student mathematics portfolio entries, and fourth-grade teachers devoted about one-third of their out-of-class time to scoring pieces from the writing portfolios.

Others have reported that finding appropriate tasks is a significant problem when portfolios are first implemented (Koretz et al., 1994b), but this does not seem to be the case in

Table 5.4

Use of Out-of-Class Portfolio-Related Preparation Time

Activity	Percentage of Preparation Time for Writing Portfolios (Fourth-Grade Teachers)	Percentage of Preparation Time for Math Portfolios (Eighth-Grade Math Teachers)
Scoring/evaluating student work	33	46
Preparing portfolio lessons	20	16
Finding appropriate tasks or materials	16	18
Attending portfolio training	12	8
Discussing portfolios with colleagues	11	8
Photocopying student work	5	5
Other	2	—

Kentucky. According to their survey responses, Kentucky fourth-grade teachers and eighth-grade mathematics teachers did not devote an inordinate amount of time to finding appropriate tasks. Moreover, two-thirds of the teachers in both grade levels agreed or strongly agreed that it had become easier to find good portfolio tasks.

However, fourth-grade teachers and eighth-grade mathematics teachers reported that the overall burden of portfolios had not declined from previous years and that scoring remains the most time-consuming task. About 60 percent of the fourth-grade teachers and about 75 percent of the eighth-grade mathematics teachers disagreed or strongly disagreed that portfolios were "less of a burden" this year than previously. In particular, both fourth- and eighth-grade teachers thought they spent too much time on scoring: sixty-one percent of the fourth-grade teachers and 80 percent of the eighth-grade mathematics teachers strongly agreed that scoring was too time-consuming. Our computations based on teachers' reports of typical time-to-score and number of portfolios scored suggest that fourth-grade teachers spent approximately 15 hours during the year scoring writing portfolios, while eighth-grade mathematics teachers spent approximately 23 hours during the year scoring mathematics portfolios.

Portfolio scoring is time-consuming; fourth-grade teachers and eighth-grade mathematics teachers typically required 30 minutes to score an average writing or mathematics assessment portfolio. An easy portfolio required about 15 minutes to score, and a difficult one required about 45 minutes. In contrast, the Kentucky Department of Education has received reports from teachers suggesting that mathematics portfolios take much longer to score than writing portfolios, perhaps four times as long.[15] There was some variation between teachers in reported time to score a typical portfolio, but the differences were relatively small. For example, 61 percent of the teachers reported that they spent 20 to 30 minutes scoring a typical portfolio; only 16 percent spent 45 minutes or more scoring a typical portfolio. The typical (median) fourth-grade teacher scored 31 portfolios in 1994–95, while the typical eighth-grade mathematics teacher scored about 47.

SCORING CRITERIA

Eighth-grade mathematics teachers reported some problems applying the scoring criteria.[16] The most difficult criterion to apply was "understanding/connecting core concepts." Over one-third of the eighth-grade mathematics teachers reported problems with this criterion on more than one-half of the portfolios they scored, and 60 percent had difficulty applying this criterion on at least one-quarter of their students' portfolios. The other criteria presented problems less frequently, but they still were difficult for teachers to apply. Twenty-two to thirty-eight percent of the eighth-grade mathematics teachers had problems scoring one-quarter or more of the assessment portfolios on problem-solving, reasoning, or mathematical communication. Similarly, 35 percent of the eighth-grade

[15]E. Riedy, personal communication, February 26, 1996.

[16]This question was not included in the fourth-grade survey.

mathematics teachers had problems assigning an overall score to one-quarter or more of their mathematics assessment portfolios.

INSTRUCTIONAL IMPACT

Fourth-grade teachers and eighth-grade mathematics teachers reported that portfolios led to innovations and to shifts in curricular emphasis. About 60 percent of the teachers in both grade levels either somewhat agreed or strongly agreed that portfolios led them "to be more innovative in planning" lessons and activities. Teachers in both grades were in greatest agreement that portfolios made it difficult to cover the regular curriculum. Overall, two-thirds of the teachers strongly agreed with this statement, and 90 percent agreed at least somewhat. One of the consequences of this would appear to be a reduction in time devoted to certain aspects of writing and mathematics. Two-thirds of fourth-grade teachers agreed or strongly agreed that portfolios caused them to deemphasize the mechanics of writing. Similarly, over 80 percent of the eighth-grade mathematics teachers agreed or strongly agreed that portfolios caused them to shift emphasis to writing and problem-solving and away from computation and algorithms. Fourth-grade teachers indicated they had taken time away from, in order: special projects; instruction on spelling, punctuation, and vocabulary; instruction in science or social studies; reading instruction; and instruction in mathematics.

Another way to look at the degree of innovation is to consider the nature of the tasks that are assigned. On the one hand, in eighth grade most tasks were drawn from the regular mathematics curriculum. More than one-half of the eighth-grade mathematics teachers frequently assigned tasks that only required skills from the regular mathematics curriculum, and about one-third frequently assigned tasks that introduced regular topics not yet covered. On the other hand, about two-thirds of the eighth-grade mathematics teachers frequently assigned tasks that extended the regular curriculum within the domain of mathematics. About one-third of the teachers frequently assigned tasks that were more difficult or challenging and tasks that extended the regular curriculum into other subjects, and about 20 percent frequently assigned tasks that were novel or not related to the curriculum. Furthermore, tasks classified by our respondents as novel and challenging were assigned at least occasionally by about 90 percent of the eighth-grade mathematics teachers.[17] When teachers assigned novel or difficult tasks, they usually provided some sort of assistance. Two-thirds of the eighth-grade mathematics teachers frequently gave assistance in the form of hints after the students had struggled with the problem or by preparing students ahead of time by teaching them the skills they would need. About one-half of the teachers frequently assigned similar but simpler problems to help students deal with novel or challenging tasks.

Finally, the surveys provided some indirect information about the effects of the portfolios on students. Although eighth-grade teachers reported that students were working on more novel and extended tasks, they also reported that students had negative reactions to certain aspects of the portfolios. Three-quarters of the eighth-grade mathematics teachers

[17]This question was not included in the fourth-grade survey.

strongly agreed that the emphasis on writing led students to become tired of writing. Two-thirds strongly disagreed with the statement that students enjoyed portfolio tasks more than other mathematics assignments. These reactions might be due to a number of factors, including the unexpected introduction of writing into the mathematics curriculum, the choice of unnecessarily textual tasks by teachers, or an imbalance in the relative emphasis placed on literacy skills and mathematical skills. However, further research is necessary before any conclusions about student perceptions of portfolios can be drawn with confidence.

6. PREPARING STUDENTS FOR THE KIRIS ASSESSMENTS

The success of KERA hinges substantially on the approaches teachers follow to prepare students for the KIRIS assessments. Some methods of preparation will improve instruction; others may leave instruction unimproved or even degrade it. Similarly, the validity of gains in KIRIS scores will depend on the methods used to prepare students for the assessment. Some methods may produce real, generalizable gains in student performance, while others will inflate scores and create an illusion of progress. In addition, inappropriate administration of the assessment may distort gains in scores.

We therefore asked principals and teachers a variety of questions pertaining to test preparation and administration of the assessment. Test preparation shades into instruction, and one goal of the current education reform movement is to diminish further the distinction between them. Our questions therefore spanned the range from generalized instructional changes (e.g., giving more homework, raising expectations, placing more emphasis on higher-order thinking skills, etc.) to methods that are tightly tied to the assessment itself (e.g., practicing old KIRIS items). The methods about which we asked also ranged from clearly legitimate to clearly illegitimate (e.g., providing hints on correct answers while administering KIRIS).

Both principals and teachers reported widespread reliance on a variety of approaches to preparing for KIRIS, including setting higher expectations, placing more emphasis on higher-order thinking skills, attempting to improve students' motivation to do well on the assessment, using practice tests, and giving instruction on test-taking skills. However, few teachers reported focusing a great deal on more homework, and only about one in four reported focusing a great deal on more or harder work in school. An appreciable minority of teachers reported that inappropriate test-administration practices occurred at least occasionally in their schools.

Educators' explanations of the effectiveness of these approaches in raising KIRIS scores showed a disturbing pattern. Despite the widespread emphasis on broad instructional changes, only a minority of teachers suggested that broad improvements in knowledge and skills, or even improvements in knowledge and skills emphasized in KIRIS, contributed a great deal to score gains in their schools. Far more educators gave credit to factors that have the potential to inflate test scores—in particular, increased familiarity with KIRIS and the use of practice tests and other test-preparation materials.[18] Although these findings may stem in substantial part from the newness of the assessment and may not predict responses over the longer term, they warrant concern. In combination with responses to other questions in the surveys, these findings suggest the need for further research to assess more

[18]We use the term "inflate" to indicate increases in test scores that are not generalizable, i.e., that are not accompanied by increases in the skills and abilities that the assessment is intended to measure.

directly the meaningfulness and generalizability of both short- and longer-term gains on KIRIS.

METHODS FOR PREPARING STUDENTS

Many of the methods for preparing students for an assessment fall on a continuum from instruction focused broadly on the assessment's domains to test preparation focused narrowly on the content and format of the assessment. For purposes of description, however, we have classified them into four categories. *Instructional approaches* include activities designed to teach students the underlying skills and abilities assessed by KIRIS. *Direct test preparation* focuses more narrowly on the specifics of the test and includes the use of practice tests and similar materials and instruction in test-taking skills. *Motivational approaches* include methods to encourage students to try hard, either on the assessment or in school more generally. *Questionable test preparation and administration* includes methods that are often considered inappropriate or even unethical, such as providing practice on secure (nonreleased) test items and providing hints during testing. The distinctions between these categories, however, are not always precise; in particular, instructional approaches and direct test preparation overlap, and the division between them is arguable.

Instructional Approaches

In the view of principals, broad instructional changes were among the most prevalent responses to KIRIS. (These instructional changes are discussed in more detail in the preceding section, but several are noted here to place the narrower forms of test preparation noted below into context.) Three-fourths or more of principals reported encouraging their teachers "a great deal" to raise expectations for students at every level of achievement. Seventy-seven percent reported giving their teachers a great deal of encouragement to focus instruction more on higher-order thinking skills, and even more (86 percent) reported giving teachers a great deal of encouragement to "improve instruction generally." A somewhat smaller majority of principals (58 percent) said their schools' emphasis on higher-order thinking skills had actually increased greatly.

Fewer teachers than principals reported a strong emphasis on broad improvements in instruction, but nonetheless a solid majority (58 percent) reported that they had focused a great deal on "improving instruction generally" in their efforts to improve scores on KIRIS. (Almost all teachers reported focusing more than a small amount—that is, "a moderate amount" or "a great deal"—on general improvements in instruction.) Only 29 percent of teachers, however, reported focusing a great deal on requiring more or harder work in school, and only 9 percent reported focusing a great deal on more homework.

A long-standing concern about assessment-based accountability is the potential for instruction tailored to raise scores rather than improve mastery more generally. We approached this issue with questions asking principals and teachers about their efforts to align instruction with the test and about changes in emphasis on untested material. Alignment of instruction with the assessment is an ambiguous category of test preparation. Some degree of alignment is one of the primary goals of the program; it is seen as one of the

most important tools for focusing attention on desired outcomes. Nonetheless, excessive alignment can inflate scores by narrowing instruction to focus on topics or types of tasks emphasized by the assessment at the expense of other important aspects of the broad domain of knowledge the assessment is designed to measure.

Kentucky principals reported efforts to align instruction with KIRIS about as frequently as the more general instructional changes just noted. Seventy percent of principals reported encouraging their teachers a great deal to focus instruction on "skills or content likely to be on KIRIS" (in comparison with the 77 percent who offered similar encouragement to increase the focus on higher-order thinking skills). About half of them (46 percent) reported that their schools' emphasis on material likely to be on KIRIS had increased a great deal (compared with the 58 percent that reported a great increase in emphasis on higher-order thinking skills), and nearly all reported at least a moderate increase in emphasis on material likely to be on the assessment.

Teachers were slightly less likely to report a strong emphasis on aligning instruction with KIRIS than on general improvements in instruction. Forty percent of teachers reported focusing a great deal on "increasing the match between the content of instruction and the content of KIRIS" in their efforts to raise scores—somewhat fewer than the 58 percent who reported focusing a great deal on general instructional improvements. About half of the teachers (48 percent) reported focusing a great deal on using "KIRIS-like tasks" in regular instruction.

Despite these efforts at aligning instruction with KIRIS, only a minority of principals reported lessened emphasis on untested material. About a third of principals reported that their school's emphasis on important aspects of the pre-KIRIS curriculum had decreased somewhat or greatly, and a roughly comparable number reported a decrease in emphasis on material that is unlikely to be tested even though it is in a tested subject area. (Only 6 percent and 5 percent, respectively, reported that emphasis on this material had decreased greatly.) Nearly half of the principals (44 percent) reported a decrease in emphasis on untested subject areas.

In contrast, most teachers did report a reduced emphasis on untested material. Because fourth-grade teachers teach many subject areas, they were asked to express their agreement or disagreement with the statement that "KIRIS has caused some teachers to deemphasize or neglect untested subject areas."[19] Eighty-seven percent agreed (40 percent strongly). Eighth-grade mathematics teachers were asked a narrower question: whether KIRIS had caused some mathematics teachers to deemphasize or neglect untested mathematics topics. Their responses, however, were similar: eighty-six percent agreed (43 percent strongly) that such changes had occurred. As noted earlier, teachers also reported deemphasizing some aspects of instruction specifically in response to the portfolio component of the assessment: the mechanics of writing, in the case of many fourth-grade teachers, and computation and algorithms, in the case of eighth-grade mathematics teachers.

---

[19]To balance the survey, some of the questions that required respondents to express agreement or disagreement with a statement expressed a negative conclusion about the assessment, while others expressed a positive conclusion.

In the eyes of teachers, the rewards and sanctions associated with KIRIS contributed to the deemphasis on untested material, and material deemphasized included some that is important. Teachers were asked to express agreement or disagreement with the statement that "imposing rewards and sanctions based on KIRIS scores . . . causes teachers to ignore important aspects of the curriculum." Sixty-seven percent of fourth-grade teachers and 82 percent of eighth-grade mathematics teachers agreed with this statement; roughly 40 percent strongly agreed.

Motivational Approaches

Kentucky educators reported substantial reliance on motivational approaches to raising KIRIS scores, although slightly fewer teachers than principals reported reliance on them. Eighty-one percent of principals reported encouraging teachers a great deal to improve students' motivation to do well on KIRIS. Sixty-five percent of teachers reported that they focused a great deal specifically on student motivation to do well on KIRIS in their efforts to raise scores, and about the same number (58 percent) reported focusing a great deal on student motivation generally.

Teachers were asked about the extent to which their schools relied on each of eight specific incentives to encourage students to do well on KIRIS. Virtually all reported that their schools placed some reliance on discussing the importance of good performance on KIRIS to the school, and over half (58 percent) reported placing a great deal of reliance on this approach. Between 80 percent and 90 percent of teachers reported that their schools placed at least some reliance on each of four other motivational approaches: (1) praising or criticizing performance on practice tests, (2) reporting student performance to parents, (3) including KIRIS scores or work from assessments in students' permanent records, and (4) scoring and providing feedback to students on the current year's common open-response items. Between 35 percent and 43 percent of teachers reported that their schools place a great deal of reliance on each of these approaches.

Fewer teachers reported that their schools prompted students to do their best by using KIRIS performance for grades or promotion or placement decisions. About half reported that their schools placed some reliance on using teacher-assigned scores on the current year's common open-response items in report card grades; one-third reported some reliance on using KIRIS results for placement in special programs; and about one-fourth reported some reliance on the use of KIRIS results in making decisions about promotion. Seventeen percent reported that their schools relied a great deal on counting scores in report card grades, and very few reported a great deal of reliance on using KIRIS results in promotion or placement decisions.

Social recognition and rewards were used in many schools. Two-thirds of teachers reported some reliance on public recognition for performance, and a similar number reported some use of special activities and prizes. More than one-fourth reported a great deal of reliance on these approaches.

Direct Test Preparation

We also questioned principals and teachers about direct test preparation, such as giving students instruction on test-taking skills or practice on old assessment items or practice tests. One set of questions asked principals how much they encouraged teachers to rely on these techniques—e.g., "How much have you encouraged your teachers to use old KIRIS items, scrimmage tests, or other test-preparation materials [in response to KIRIS]?" Teachers were asked questions about their reliance on these techniques specifically as a means of raising scores—e.g., "In trying to improve scores on the KIRIS assessments, how much have you focused on using practice tests and other test-preparation materials?" These practices, like curriculum alignment, are inherently ambiguous. A certain amount of practice can be desirable—for example, to provide students with the necessary degree of familiarity with test formats and to illustrate concretely the types of knowledge and skills that are expected of them. For these reasons, KDE has encouraged teachers to use KIRIS items as practice for KIRIS, and the scrimmage tests are designed expressly for this purpose. Here again, however, excessive reliance on direct test preparation runs the risk of inflating scores (and siphoning limited instructional time away from other activities).

Most principals reported placing a great deal of emphasis on both test-preparation materials and test-taking skills, especially the former. Eighty-two percent of principals reported encouraging their teachers a great deal to use test-preparation materials. Somewhat fewer of the principals—66 percent—reported encouraging teachers a great deal to teach test-taking skills. All principals, however, reported encouraging teachers at least somewhat to use both of these approaches.

Teachers were less likely to report heavy reliance on these techniques than principals were to report greatly encouraging their use, but nonetheless, most teachers reported substantial reliance on them. Almost all teachers (92 percent) reported focusing at least a moderate amount (that is, more than "a small amount") on test-taking skills, and 48 percent reported focusing a great deal on them. Reported reliance on practice tests and other test-preparation materials was somewhat less widespread. Seventy-seven percent of teachers reported focusing at least a moderate amount on practice tests and test-preparation materials, and 36 percent reported focusing a great deal on them. Moreover, virtually all teachers (98 percent) reported that students in their schools were given practice on the previous year's KIRIS common items at least occasionally, and about half reported that students were frequently given practice on them. Slightly more fourth-grade teachers (58 percent) than eighth-grade mathematics teachers (45 percent) reported that such practice was frequent. Most teachers (83 percent) reported that students in their schools are at least occasionally provided practice on items that are highly similar to the previous year's matrix items, and 40 percent reported that such practice is frequent.[20]

[20]These last two questions overlap, in that the previous year's common items are in principle similar to the matrix items. However, the slightly smaller percentages replying in the affirmative to the question about practice with items highly similar to the previous year's matrix items suggest that some teachers interpreted it to refer to practice items other than old common items.

In our mail surveys, we also asked teachers how much time they devoted to six specific types of test-preparation activities during the previous school year. Eighth-grade teachers were asked how many partial or full class periods they devoted to each; fourth-grade teachers were asked the number of days during which they did each. All teachers were also asked to convert these answers to a number of minutes in the entire year.[21] We converted these to hours and to class-period-equivalents, assuming 50 minutes per class.

Fourth-grade teachers' responses varied greatly: Some reported very little use of any of the test-preparation activities about which we asked, while others reported allocating considerable time to them. Among fourth-grade teachers as a group, more time was allocated to the use of released KIRIS items than to any of the other forms of practice tests about which we asked. The typical (median) teacher reported allocating 6.7 hours over 20 days on sample or released KIRIS items (Table 6.1), but a fourth of the teachers reported 2.7 hours or less, while another fourth reported 15 hours or more. (In Tables 6.1 and 6.2, the percentiles refer to a ranking of teachers based on the time they report allocating to each activity. Thus, for example, Table 6.1 shows that 75 percent of teachers reported allocating 15 hours or less to the use of released KIRIS items.) The typical teacher reported allocating 2.5 hours to practice tests that he or she had developed independently in preparation for KIRIS, while more than a fourth reported no use of such tests, and another fourth reported 10 hours or more. The typical teacher did not use either scrimmage tests (provided by the Kentucky Department of Education) or district- or school-developed practice tests and reported spending less than an hour using student work from previous KIRIS assessments.

Despite the fact that about half of the teachers reported focusing a great deal on test-taking skills, fourth-grade teachers generally reported allocating much less time to instruction in test-taking strategies than to the use of practice test items. The typical teacher reported allocating less than 2.5 hours to test-taking skills, and 75 percent reported 5 hours or less. It is possible, however, that they focused on test-taking skills in the context of other activities (such as using practice tests) the primary focus of which was not test-taking skills per se.

The total time allocated to all of these forms of practice tests together is of course considerably larger than the time spent on any one, but it cannot be estimated precisely from the survey. The median teacher reported allocating 15 hours in total to all five of the types of practice tests about which we asked (excluding instruction in test-taking skills), and a fourth of the teachers reported 25.5 hours or more. These totals may overstate their allocation of time to practice tests, however, because it is possible that some teachers reported the same time more than once. (For example, a teacher who gave a practice test including both KIRIS and self-developed items might have mistakenly reported the total time for that practice test for both categories.) These totals, even if overstated somewhat, represent a sizable amount of time but only a modest share of the total instructional time available during the year. For

[21]For example, the question asked of eighth-grade teachers read as follows: "Last year (1993–94), how much did you use each of the following methods to prepare students for KIRIS? EXAMPLE: if you devoted about 5 minutes per day for approximately 10 days, you would enter 10 under 'Partial or full class periods' and 50 under 'Total minutes in entire year.'"

Table 6.1

Reported Hours Devoted to Test-Preparation Activities, Fourth-Grade Teachers

Activity	25th Percentile	50th Percentile	75th Percentile
Used released KIRIS items	2.7	6.7	15.0
Used student work from previous KIRIS assessments	0	0.8	3.0
Used scrimmage tests	0	0	2.0
Used district or school practice tests	0	0	3.0
Used own practice tests	0	2.5	10.0
Taught test-taking strategies	1.3	2.5	5.0

example, if one assumes that a typical fourth-grade teacher has available 825 instructional hours per year (5 hours per day for 165 days), the 25.5 hours reported by the teacher at the 75th percentile represents roughly 3.1 percent of available instructional time.[22]

Eighth-grade mathematics teachers reported spending considerably fewer hours but a larger share of their instructional time on these test-preparation activities. As in fourth grade, released KIRIS items were allocated the most time: 3.3 hours by the median teacher, and 6.7 hours or more by 25 percent of the teachers (Table 6.2). The typical eighth-grade mathematics teacher reported using released KIRIS items during 10 class periods. Teachers' own practice tests again were given the second most time: a bit under 2 hours by the typical teacher, and 5 hours or more by 25 percent of the teachers. Little time was reportedly allocated to the other types of practice tests about which we asked. As those teaching fourth grade, teachers reported allocating relatively little time to instruction in test-taking skills—an hour in the case of the median eighth-grade mathematics teacher.

Table 6.2

Reported Hours Devoted to Test-Preparation Activities, Eighth-Grade Mathematics Teachers

Activity	25th Percentile	50th Percentile	75th Percentile
Used released KIRIS items	1.5	3.3	6.7
Used student work from previous KIRIS assessments	0	0.8	1.7
Used scrimmage tests	0	0	0.8
Used district or school practice tests	0	0	1.8
Used own practice tests	0	1.7	5.0
Taught test-taking strategies	0.8	1.0	2.5

[22]This assumes that a total of 10 days of a typical 175-day year are spent on noninstructional activities such as testing, field trips, half-days off for professional time, etc.

The greater reported allocation of time to test preparation by eighth-grade mathematics teachers becomes apparent when these results are compared with available instructional time. The median eighth-grade mathematics teacher reported allocating 6.5 hours to the five types of practice tests together. This corresponds to about eight 50-minute class periods, or perhaps 4.7 percent of annual instructional time.[23] A fourth of teachers reported more than 15 hours, corresponding to 18 class periods, or nearly 11 percent of instructional time. Even allowing for some overstatement from double-counting of time, as explained above, this suggests that many eighth-grade mathematics teachers are allocating an appreciable share of available instructional time to practice tests.

The large majority of teachers (74 percent in grade four and 66 percent in grade eight) reported that they did most of this test-preparation activity throughout the year. Only about one-fourth of the teachers reported that they conducted it mostly in the month preceding KIRIS, and only a handful reported doing it mostly during the two weeks before the assessment.

Even though teachers reported allocating appreciable time to practice tests, it may be surprising that they do not allocate more, given that they are urged to use old KIRIS items and scrimmage tests to prepare students for KIRIS. On the other hand, it is possible that activities that many observers would consider test preparation, such as the use of tasks highly similar to KIRIS tasks, might not have been included in some teachers' responses, given that our surveys asked only about very specific types of practice tests.

Whether this allocation of time to test preparation is on balance desirable or undesirable cannot be determined from our survey responses. Additional information about the specific activities undertaken would help determine the extent to which they constitute desirable instruction, and information on the activities they displace would help determine whether their *net* effect on instruction is desirable. Empirical data on the generalizability of students' gains on KIRIS would be needed to reach a firm conclusion about the impact on achievement of all of the activities undertaken to prepare students for KIRIS; gains that do not generalize might suggest that some of the direct test preparation may be contributing to score inflation.

Questionable Test Preparation and Administration

We asked teachers to comment on the frequency with which a number of questionable test-preparation activities or test-administration practices occurred in their schools in preparing students for the open-response items on the KIRIS transitional assessments. Because we feared that teachers would find it difficult to provide us with information about questionable test-preparation activities and inappropriate test administration, we took several steps to make our questions about these topics (which were asked by telephone) less threatening. Teachers were asked about practices throughout their schools, rather than their own practices. (For example, one question was "To the best of your knowledge, how frequently does each of the following practices occur in your school in preparing for the open-

[23]This assumes 8,250 instructional minutes per year (165 50-minute class periods).

response items on the transitional assessment?") Respondents were told that people disagree about which practices are desirable. They were reminded that their responses were confidential and that our intent was to describe practices throughout the state, not to judge individual schools. Nonetheless, pilot interviews suggested that the questions made some respondents uncomfortable. For example, even though the questions were asked about the school as a whole, one pilot respondent paused after one of the questions and then responded several times that *she* had never engaged in the practices in question. These factors raise the risk of "social-desirability bias" in the results, which in this case could entail underreporting the actual incidence of the activities in question. We have no direct evidence, however, that underreporting occurred.

The responses to these questions were mixed. Only a minority of teachers reported any instances in their schools of any of the questionable practices about which we asked, and the frequency of these practices may be low even in the schools in which they are reported. Nonetheless, appreciable percentages of teachers reported some of the practices.

Few teachers reported misuse of secure testing materials. Certain matrix items in each assessment are secure and should not be retained or used in preparing students for subsequent assessments. Only a handful of teachers reported knowing of any instances in which those rules were violated: six percent reported that someone had obtained nonreleased items, and 4 percent reported that someone had obtained student responses to nonreleased items (Table 6.3).[24]

Appreciable minorities of teachers, however, reported some incidence of other inappropriate test-administration activities about which we asked. (Recall that these were asked specifically about the open-response items; some of these practices might be appropriate in the portfolio assessment program.) More than a third of teachers reported that questions had been rephrased during the administration of the assessment, and 12 percent report that this was frequent (Table 6.3). One teacher in five reported that staff in their schools at least occasionally answered questions about the content of the assessment during testing, and the same percentage reported that revisions were recommended at least occasionally either during or after testing. Seventeen percent reported that hints on correct answers were at least occasionally provided. Relatively few (9 percent) reported that at least occasionally, answers were edited or changed.[25]

---

[24]In response to another question that did not use the word "nonreleased," 19 percent of teachers reported that students were frequently "given practice on the previous year's matrix-sampled items," and 61 percent reported that this was done at least occasionally. These results, however, seem to contradict the finding that only 6 percent reported that anyone in the school had even obtained the previous year's "nonreleased" matrix items. Analysis of pilot interviews suggested that because of the omission of the word "nonreleased" in the former item, many teachers may have failed to understand that the question was intended to refer only to nonreleased items, despite the explicit mention of matrix items. (We therefore did not include the results from this question in Table 6.3.)

[25]In similar surveys in Maryland, we found that somewhat fewer teachers reported these practices in the administration of the Maryland State Performance Assessment Program (MSPAP) assessment, a difference that might stem from the lower stakes in Maryland. For example, in Maryland, 27 percent (in contrast to 36 percent in Kentucky) reported rephrasing of questions; 13 percent (versus 21 percent) reported answering questions about content; 14 percent (versus 21 percent)

Table 6.3

Percentage of Teachers Reporting Incidence of Questionable Test-Preparation and Administration Practices

Practice	Yes	Occasionally or Frequently	Frequently
Obtained last year's nonreleased matrix items*	6		
Obtained student responses to last year's matrix items*	4		
Questions rephrased during testing time		36	12
Questions about content of assessment answered during testing time		21	6
Revisions recommended during or after assessment		21	6
Hints provided on correct answers		17	3
Changes or edits made to answers in assessment booklets		9	2
Items read for students		*43*	*2*
Responses written for students		*15*	*4*

NOTE: Items marked with an asterisk allowed only "yes" or "no" answers. In all other cases, respondents were allowed "never," "occasionally," or "frequently." Italicized items may pertain to students with special needs; see explanation in the text.

Teachers were also asked about two practices that would be appropriate in some cases but inappropriate in most: reading KIRIS items for students and writing answers for them. (These items are italicized and are shown below the last separator in Table 6.3.) Nearly half (43 percent) reported that items were at least occasionally read, while 15 percent reported that answers were at least occasionally written. These findings are difficult to interpret without additional information about the contexts in which this occurred because there are instances under which these practices are accepted. Under Kentucky's assessment guidelines, both reading and writing for students would be appropriate in the case of students with disabilities, provided that the students' Individualized Educational Plans (IEPs) call for those accommodations in both instruction and testing. With hindsight, our surveys should have asked about the incidence of these practices for students not formally identified as disabled or for whom these accommodations are not specified in IEPs.

PERCEIVED CAUSES OF GAINS ON KIRIS

We asked both principals and teachers whether their schools' KIRIS scores had increased. Those who answered positively were asked to explain their gains by reporting their opinion about the amount each of seven factors had contributed to them.

In the aggregate, educators' responses revealed some lack of confidence in the meaningfulness of their schools' gains in scores. Although, as noted earlier, most educators reported a strong emphasis on broad instructional changes in response to KIRIS, "broad improvements in knowledge and skills" was one of two factors cited *least* frequently by

reported that revisions had been suggested; and 2 percent (versus 9 percent) reported that changes had been made to answers (Koretz et al., 1996).

teachers as having contributed a great deal to their schools' KIRIS gains, along with increases in student motivation: Only 16 percent cited each of these factors (Table 6.4). "Improvements in students' mastery of knowledge and skills that are emphasized in KIRIS" were cited by almost as few teachers (24 percent). Cited most frequently as having contributed a great deal to KIRIS gains were "increased familiarity with the KIRIS assessments" (55 percent) and "work with practice tests and other preparation materials" (51 percent). Despite teachers' reported focus on test-taking skills, improved test-taking skills were cited by fewer teachers (34 percent). In the words of one teacher, "Students are only doing better on tests . . . because the teachers are better prepared at teaching it. [It is] not that the students are any brighter. Scores have improved not because the students are more knowledgeable."

As a group, principals were somewhat more positive about their schools' gains on KIRIS: They were more likely to report that improvements in knowledge and skills had contributed a great deal to their score gains and less likely to cite improved test-taking skills. Even principals, however, were markedly more likely to cite familiarity with the assessment (56 percent) than broad improvements in knowledge and skills (31 percent) or improvements in the knowledge and skills emphasized in KIRIS (34 percent—Table 6.4).

These opinions about gains in educators' own schools were consistent with their opinions about the effects of KIRIS in Kentucky more generally. As noted earlier, the large majority of educators (87 percent of teachers and 71 percent of principals) expressed at least some agreement with the statement that some schools had found ways to improve scores without improving education.

However, most Kentucky educators reported that improvements in knowledge and skills contributed at least a "moderate amount" (i.e., more than "a small amount") to the KIRIS gains in their schools. Sixty-five percent or more of teachers and 77 percent or more of principals reported that broad improvements in knowledge and skills had contributed at least a moderate amount to their score gains. Roughly 75 percent of teachers and 85 percent of principals said the same of improvements in students' mastery of knowledge and skills that are emphasized in KIRIS.

Table 6.4

Percentage of Teachers and Principals Reporting That Each Factor Contributed "A Great Deal" to KIRIS Gains in Their Schools

	Teachers	Principals
Increased familiarity with KIRIS	55	56
Work with practice tests and preparation materials	51	43
Improved test-taking skills	34	22
Differences between cohorts	26	19
Improvements in knowledge and skills emphasized in KIRIS	24	34
Broad improvements in knowledge and skills	16	31
Increased student motivation	16	20

These responses raise concerns about the validity of initial gains in KIRIS scores. The goal of the accountability program is improvements in students' knowledge and skills. If educators are correct that factors such as work with practice tests have contributed more to gains than have improvements in knowledge and skills, gains in scores may be misleading as indicators of success. As noted, the impact of test preparation and familiarity are ambiguous, and survey data alone are not sufficient to test whether scores are inflated. Moreover, familiarization is likely to contribute substantially to initial gains after any new assessment is introduced. Nonetheless, these results—particularly in the context of the finding that initial KIRIS gains were not echoed in scores on either the National Assessment of Educational Progress or the American College Testing college-admissions tests (Hambleton et al., 1995)—strongly suggest the need for further investigation of the meaningfulness of both short- and longer-term gains on KIRIS. These issues are discussed in more detail in the following section.

7. DISCUSSION

The views of fourth-grade teachers, eighth-grade mathematics teachers, and elementary and middle-school principals paint a complex portrait of KIRIS in the 1994–95 school year, suggesting both successes and problems. Some key findings are recapitulated here, and a discussion of their implications follows.

SUMMARY OF KEY FINDINGS

A majority of principals and eighth-grade mathematics teachers (about 60 percent) voiced global support for the program; fourth-grade teachers were almost evenly split between supporters and opponents. About half of principals reported becoming more positive toward KIRIS over the preceding years, while about a fourth reported becoming more negative. About three-fourths of the principals said KIRIS imposed more than a minor burden on their schools. Some of the burdens are intended, however, and about two-thirds of those principals reported that the benefits of the program balanced or exceeded the burdens it imposed. Moreover, about two-thirds of principals reported that the program had become easier to accommodate in the several years it had been in place.

Teachers, however, reported that KIRIS has caused high stress. Most teachers strongly agreed that KIRIS has put teachers under "undue" pressure. Most teachers reported that teacher morale in their schools is low and has been harmed by KIRIS, and about half reported that KIRIS has reduced their own job satisfaction. A sizable minority reported that KIRIS has also decreased the morale of their students.

Teachers were roughly evenly divided with respect to a fundamental tenet of the program: that all students can learn to a high level. Interestingly, most teachers agreed that this is the right message to give to Kentucky students regardless of its feasibility. About two-thirds of teachers and principals agreed that the current improvement threshold for their schools is realistic, but very few considered the long-term goal of reaching an accountability index of 100 to be realistic. Support for the accountability component of the program was low; only about a quarter said they support the imposition of rewards and sanctions based on KIRIS.

Educators voiced both positive and negative views of the KIRIS assessment per se. On the positive side, about three-fourths of teachers reported that KIRIS tests a wider range of skills than do multiple-choice tests and that KIRIS tasks are based on realistic situations. The percentages reporting that the student achievement information yielded by KIRIS is accurate varied dramatically, from 52 percent to 81 percent, with the multiple-choice items rated favorably most often and performance events and portfolios least often. Principals' and teachers' views of the reasonableness of the KIRIS components for drawing conclusions about school effectiveness were similar. On the negative side, about half of the teachers strongly agreed that scoring standards for KIRIS are inconsistent over time, and a similar percentage strongly agreed that the curriculum content for the assessments is not defined well enough

for them to prepare students adequately. Over 60 percent of principals and teachers strongly agreed that schools with highly transient populations are at an unfair disadvantage on KIRIS. About half of the teachers reported that the emphasis on writing in KIRIS makes it hard to judge the mathematics achievement of some students.

The central goal of KIRIS is to improve instruction, and many Kentucky educators believe it has increased educators' efforts in this regard and has caused changes in instruction. All principals reported that they have encouraged their teachers to improve instruction in response to KIRIS. In addition, about three-fourths of principals reported that KIRIS has been a useful tool for encouraging instructional change by teachers who are very resistant to changing, and a majority of teachers concurred that the program has encouraged some such teachers to change their instruction. Almost all teachers reported that they had focused a moderate amount or a great deal on improving instruction in their efforts to raise scores.

Teachers reported that KIRIS has produced both positive and negative instructional effects, but somewhat more reported positive effects. Most teachers reported that at least one part of the KIRIS assessments had had a moderate or great deal of positive effect on instruction, but a majority of teachers also reported that at least one part of the assessment had had a moderate or great deal of negative effect on instruction.

Four-fifths or more of teachers reported increasing their emphasis on writing (in fourth grade), problem-solving, and communication of mathematics. Teachers' responses to open-ended questions about changes in instruction, while harder to quantify, were largely consistent with these results. The positive comments made by the most teachers concerned writing—in particular, noting their own increased emphasis on writing or improvements in students' writing and communication skills. A sizable number also commented that KIRIS had led to more focus on problem-solving and thinking skills. Some teachers also commented that KIRIS had led teachers to focus more on real-life applications, hands-on activities, and cooperative learning.

The perceived negative instructional effects cited by teachers were often related to the amount of time taken away from other instruction to prepare for or administer the assessment. Some teachers reported deemphasis on language mechanics, number facts, computation, and mathematical algorithms. Teachers also expressed negative comments about the amount of time students spend writing, and some commented that students are "burnt out" by all of the writing they now need to do.

Educators' responses to questions about the portfolio program were particularly striking in pointing to both successes and problems. Many teachers reported that the portfolio program had led them to be more innovative in planning and teaching. Substantial majorities of teachers indicated that portfolios had had a moderate amount or a great deal of positive effect on instruction in their schools, and in fourth grade (where teachers had more experience with portfolios), portfolios were cited by about half of the teachers as having had a great deal of positive impact. However, a smaller majority of teachers reported that the portfolio assessment had had more than a small *negative* effect on instruction, and about a fourth of teachers said that the portfolio assessment had had a great deal of negative impact.

(Some teachers reported both positive and negative effects.) Teachers reported that portfolios are time-consuming, and they reported variations in portfolio practices that have the potential to undermine the validity of comparisons among schools.

About two-thirds of teachers reported that expectations for students have changed because of KIRIS. Most of the teachers who reported a change in emphasis on high standards reported that it had been helpful to students. The responses suggest, however, that these effects on expectations may have been less substantial for students with low levels of achievement or disabilities. For example, 24 percent of teachers reported that expectations had increased greatly for high-achieving students, in comparison to 12 percent for special-education students and 16 percent for low-achieving students. In fourth grade (which was more extreme in this respect than eighth grade), 86 percent of teachers who reported a change believed that it had been helpful to high-achieving students, in comparison to 44 percent and 56 percent who said the same of special-education and low-achieving students.

Educators' responses to several of the survey's questions raised the issue of potentially inflated gains on KIRIS during the program's first years. Several of the responses indicated substantial direct test preparation, and while these activities are in some measure encouraged by KDE and are ambiguous in terms of their effects on both scores and instructional quality, they point to a potential problem that should be explored further. Most teachers (nearly 90 percent) agreed that KIRIS has caused teachers to deemphasize or neglect untested subject areas. Almost all teachers reported relying at least a moderate amount on instruction in test-taking skills in their efforts to raise scores, and about three-fourths said the same of practice tests and related test-preparation materials. About half reported that students were frequently given practice on the previous year's KIRIS items. Substantial numbers of teachers, particularly eighth-grade mathematics teachers, reported allocating substantial amounts of instructional time to practice tests. A second set of responses points more clearly to the possibility of distorted scores. A large majority of educators (particularly teachers) agreed to some degree with the statement that some schools had found ways to improve scores without improving education. Perhaps most striking are educators' explanations of the KIRIS gains in their own schools during the first years of the program. Half of the teachers reported that familiarity with KIRIS and work with practice tests and other preparation materials had contributed a great deal to their KIRIS gains, while only 16 percent said that broad improvements in knowledge and skills had contributed a great deal, and about a fourth said the same of improvements in the knowledge and skills emphasized by KIRIS. Principals were more optimistic in their explanations, but even they were more likely to attribute their schools' gains to familiarity and test preparation. Finally, an appreciably minority of principals reported moving teachers among grades to place the more able teachers in accountability grades.

IMPLICATIONS

KERA is a sweeping reform, and even the KIRIS assessment and accountability component of KERA is an extremely ambitious undertaking that calls for large and pervasive

changes in practice. Moreover, one of the mechanisms by which the system attempts to bring about change is to use clear performance standards and substantial rewards and sanctions to pressure individuals and systems to change. Accordingly, it is not surprising that these survey results indicate a mix of favorable and unfavorable responses by educators and suggest problems as well as signs of success. A reform of this scope should be expected to produce unintended as well as intended effects. Even successful components of the program may take years to have their intended effects, as educators gradually become familiar with the system and its goals, obtain training, and learn to modify their practices in response. Moreover, initial missteps in program design and implementation are inevitable, and time will be required to discern them and to alter the program in response.

Nonetheless, the results reported have significant implications for policy. They indicate initial successes that can be built upon as well as important concerns that warrant attention by the Kentucky Department of Education in its efforts to refine the reform program. In addition, these results point to the need for additional research to monitor the program's operation, ascertain its effects, and evaluate the quality of the performance information that is the core of the KIRIS system. Surveys provide only a first look at these issues; they raise many questions that can be answered with confidence only by other kinds of research.

Lack of Support for Accountability

Particularly in the early years of a program of this sort, some degree of dissatisfaction by educators is to be expected, given the dislocations the reforms will cause and the added pressures of accountability. Indeed, some amount of dissatisfaction may signal successes. Recall that many of our respondents—both principals and teachers—reported that KIRIS has been a useful tool for getting reluctant teachers to change their practices. It is likely that many of those changes would be applauded by program advocates, and it is also likely that the teachers who reported, for example, "undue pressure" from the system include some who reluctantly bowed to pressure to make changes intended by KERA's architects. Similarly, some teachers may object to being held accountable (perhaps for the first time) for the performance of their students, regardless of the specifics of the assessment or the accountability program.

At the same time, it would be risky to discount all of the reported dissatisfaction on these grounds. "Undue pressure," for example, was reported by nearly all fourth-grade teachers, which suggests that this concern goes well beyond a subset of teachers who are reluctant to change or to be held accountable. Moreover, centralized assessment-based accountability is necessarily a blunt instrument, and it seems likely that in some instances, it may indeed create unintended and even counterproductive pressures. This is an open question that can be addressed only by further investigation.

Perceived Burdens

Principals' reports of burdens caused by KIRIS similarly could signal either success or problems (and may well indicate both). The program is designed to induce rapid

instructional change, and its architects intended to create a need for widespread retraining, so it may be a good sign that many principals report feeling pressure to effect both of those changes. On the other hand, we cannot ascertain from the survey whether these burdens are within intended bounds, and there are other reported burdens—such as time demands—that clearly are not intended. Here again, information more detailed than that provided by these surveys would help determine whether program modifications designed to reduce burden are called for.

Perceived Effects on Schooling

Many of the changes in schooling noted by principals and teachers—changes in school management, curriculum, instructional approaches, and classroom assessment, and a general raising of academic expectations—are consistent with the goals of KERA, and the fact that they were so widely reported is a very encouraging sign. To build on these changes, however, may require further investigation and actions to explore and address areas of possible concern.

One reason to follow these findings with additional investigation is that our survey questions referred to very general categories, such as "problem-solving" and "communicating mathematics." Research has shown that teachers often mean very different things by such terms and use them to refer to widely varying activities (e.g., Stecher and Mitchell, 1995). More detailed investigation would be needed to pin down the nature of these instructional changes, identify more and less desirable changes, and explore how these changes vary among teachers and schools. That information would in turn help better hone professional development efforts and other aspects of the program.

In addition, teachers' perceptions of negative instructional effects and their comments on deemphasized material warrant further exploration. It is no longer possible to get a clear baseline of pre-KIRIS instruction, so it is not feasible to obtain an unambiguous picture of the net changes in schooling induced by the program. It should be feasible, however, to obtain considerably more detailed information about effects that educators consider negative and about material that has been cut back to accommodate the demands of KIRIS, and that information could be very important in reducing unintended negative consequences of the reforms.

Some of the educational effects identified by teachers may have both good and bad elements, and their evaluation therefore may be partly a matter of judgment. For example, our surveys indicated that KIRIS has succeeded in causing a dramatic increase in the amount of writing students do in the classroom, and most observers will consider this a clear sign of success. Teachers reported not only that students spend more time writing, but also that they are better able to explain their answers. However, teachers' responses also suggest there is a negative side to this change: Many maintained that there is too much emphasis on writing and that instruction has suffered because of the amount of time consumed by students' writing. In addition, virtually all of the surveyed teachers believe that KIRIS' emphasis on writing makes it difficult to judge the mathematical competence of some students.

Further investigation is needed to explore the bases of these findings and to determine whether the tensions reported by teachers lessen as they become more adept at integrating writing (and other skills emphasized by KIRIS, such as problem-solving) into ongoing instruction focused on other content. If teachers' concerns are well-grounded and persist, the question will be raised of whether the format of the assessment places too much reliance on writing, in terms of both test validity and instructional impact. This question is a matter of policy as well as empirical evidence. With respect to validity, the question of whether the emphasis on writing obscures the mathematical competence of some students depends in part on Kentucky policymakers' definition of the domain of mathematics achievement. For example, if mathematical communication is weighted very highly in that definition relative to, say, knowledge of algebraic techniques, designing the assessment to place heavy emphasis on writing may improve the overall validity of inferences about mathematical performance even if measurement of some specific aspects of mathematics suffers. With respect to instructional impact, whether the current emphasis on writing is excessive depends in part on a policy judgment about the value of both the marginal time allocated to writing and the time taken away from other activities to make way for it. Finally, there may be trade-offs *between* impact and validity as well; policymakers may decide that the need for students to develop greater skills in writing in all subjects warrants some decrement in the validity of certain inferences based on KIRIS. The responses of teachers noted here suggest an apparent need to conduct other forms of empirical research to clarify the present trade-offs in terms of both validity and impact and to determine how well the system is meeting the intent of Kentucky policymakers.

Ranking of Assessment Components

Respondents' rankings of the impact and usefulness of the four cognitive components of KIRIS—multiple-choice items, open-response items, performance events, and portfolios—have implications for the future design of the assessment. Perhaps the most striking pattern arose in teachers' rankings of the positive effects of the various cognitive components. Recall that although the multiple-choice component was the most often cited by teachers as yielding accurate information, it was almost never cited as having a great deal of positive impact. This is consistent with the views of many advocates of performance assessment. Within the set of three performance-based components, however, teachers' ratings were in some respects at variance with the views of many performance-assessment advocates. Of the three, the one most often cited as having positive instructional effects was the open-response items, which are the least performance-oriented. These open-response items were also most often cited as very useful for improving instruction in the respondents' classrooms. Performance events, which on several dimensions are the most performance-oriented and which are justified in substantial measure in terms of anticipated positive effects on instruction, were much less often cited by teachers as having a great deal of positive effect on instruction.

If teachers' reports of instructional impact accurately reflect classroom practice, they suggest that the presumed link between instructional impact and assessment format needs empirical investigation. Although the Kentucky Department of Education deliberately chose

to use several different assessment formats for reasons of both measurement and instructional incentives, many reform advocates around the nation have assumed that the more performance-oriented a format is, the better its instructional effects are likely to be; this is linked to the notion that "good assessment mirrors good instruction." For example, many advocates prefer hands-on tasks to purely written tasks and group tasks to purely individual tasks. Reliance on these formats, however, imposes many costs, including financial costs, greater time requirements per task, increased undesirable task variance, lower reliability, and possibly threats to validity from irrelevant aspects of group composition and interaction. Teachers' responses to these surveys suggest that it may be feasible in some instances to encourage improved instruction by placing greater reliance on more traditional and less costly formats, such as essays and other open-ended written tasks. Further investigation is needed, however, to determine the accuracy of teachers' responses and, if they are accurate, to explore their causes. For example, the relative effects of the KIRIS assessment components might stem either from aspects of their design (such as format) or from other aspects of the program in which they are embedded (such as the large weight given to the open-response items in the KIRIS accountability index).

Consistent with the expectations of proponents, portfolios were often cited as having a great deal of positive impact, but they were also the only one of the four cognitive components cited by an appreciable percentage of teachers as having a great deal of negative effect. This may suggest a need for additional research that would assess *net* benefits rather than gross benefits in evaluating the impact of assessment-based reform. For example, the time spent working on a writing portfolio may be beneficial, but it may or may not be more beneficial than the activities from which time was taken to make it possible.

Specificity of Curriculum Frameworks

The fact that nearly half of the teachers strongly agreed that the curriculum framework is not well enough specified for them to prepare students for KIRIS is grounds for concern and further investigation. Teachers are being asked to adopt new forms of instruction, and it is to be expected that some will find the new directions more ambiguous than the more familiar curricula of the past. The size of the negative reaction, however, may be cause for concern. For example, in a set of similar surveys, we asked teachers in Maryland the same question about that state's new performance assessment program, the Maryland School Performance Assessment Program (MSPAP). Only 20 percent of Maryland teachers strongly agreed that the curriculum framework is not well enough specified (Koretz et al., 1996). These differences could reflect differences between the states in the specificity of their frameworks. (Kentucky teachers were probably responding in terms of KDE's early frameworks, which were markedly less specific than the current ones.) The differences might also reflect other factors, however, such as the higher stakes in Kentucky or differences between the assessments themselves.

The optimal level of specificity of the content standards and curriculum frameworks in KERA has been controversial for some time and has recently been the focus of substantial debate (e.g., Hambleton et al., 1995). Indeed, there may be no one ideal level; the optimal

may differ, for example, across subjects or grades. On the one hand, the program's designers intended for the instructional goals to be broad in order to focus educators' attention on the basic goals of the reform rather than on narrow outcomes. One of the key stakeholders once explained that KDE is specifying *what* to accomplish, not how to accomplish it, and specific curriculum decisions are "part of the how." On the other hand, if the state's framework is insufficiently detailed, teachers may respond by using the assessment itself as a surrogate for a curriculum framework. Researchers reported this phenomenon in another state using assessment as a lever for reform (Stecher and Mitchell, 1995). This in turn could increase the risk of inflated test scores if teachers focus too narrowly on the content or the format of test itself. It could also lead to inconsistent instructional change over time (if teachers draw inconsistent inferences from the assessment about the intended curriculum framework).

Over the past several years, KDE has moved to establish greater specificity of curricular expectations. The effectiveness of the new frameworks in reaching a balance between generality and specificity remains an empirical question.

Effects on Equity

A prominent theme in the current education reform movement is a desire for greater equity, both in the provision of educational opportunity and, ultimately, in educational outcomes. This theme is clearly reflected in the history and design of KERA—for example, in the fact that all schools are expected to reach the same performance standard, equivalent to having 100 percent of their students at the proficient level, within 20 years.

For this reason, the responses of educators to our questions about changing expectations for students warrant concern and further investigation. Although our findings include the good news that many teachers perceive expectations to be increasing for all groups of students, even those with low achievement or in special education, the fact that effects on expectations appeared to be somewhat more favorable for higher-achieving students raises the prospects of widening gaps in opportunities between high- and low-achieving students. Expectations are only one aspect of equity, and teachers perceptions of changes in expectations may not accurately mirror actual changes in educational practice and opportunity. Indeed, changes in KIRIS scores show a rapid decline in the proportion of students classified as novice, which may indicate improved opportunities for low-achieving students. Nonetheless, the results reported here are sufficient to indicate a need for further investigation to determine both the extent of achievement-related differences in expectations and the factors that appear to contribute to them. In addition, research is needed to explore the effects of KIRIS on the many other aspects of educational equity.

In similar surveys in Maryland, we found remarkably similar responses to the same questions about expectations (Koretz et al., 1996). This suggests that the roots of this pattern may lie in elements that the two programs share. They are quite different in terms of accountability (MSPAP entails no financial rewards and results in sanctions for only a very small number of schools), but they are similar in establishing a lowest acceptable performance standard that is very high relative to the current performance of low-achieving students. One might speculate that this leads to a large increase in expectations primarily

for students who are within striking distance of those standards—i.e., currently high-achieving students.

The Need to Explore the Validity of Score Gains

It is critically important, both for program improvement and as a matter of public accountability, to investigate the validity of score gains, particularly in the second and subsequent bienniums. KIRIS, like most assessments of achievement, is intended to represent students' mastery of broad domains of knowledge and skills, and gains on KIRIS are valuable only to the extent that they signal improved mastery of those domains. Such improvements would not be specific to KIRIS, but would generalize to an appreciable degree to other assessments developed to similar test specifications. If a sizable portion of KIRIS gains were limited to the specific assessment (for example, as a result of narrowly focused test preparation), the validity of the most important inferences based on KIRIS would be undermined. Educators, policymakers, and members of the public who drew the inference that the gains reflect improved outcomes would be misled, and schools' responses to the reforms could be misdirected.

The validity of score gains is a pressing question in the case of any high-stakes testing program because of the potential for inflated scores. Investigation of this question nonetheless has been rare until recently, but as policymakers and others increasingly become aware of its importance, it is gradually becoming more common. Research has documented that excessive test preparation and severe inflation of scores sometimes results from using traditional tests for accountability (e.g., Koretz et al., 1991; Shepard & Dougherty, 1991). Indeed, that risk is now widely accepted and is one reason many reformers currently advocate replacing traditional tests with performance assessments. (Another reason is the perceived negative effects on instruction of coaching for multiple-choice tests.) Some observers have suggested, however, that whatever their effects on the quality of instruction, performance assessments used for accountability are likely to be vulnerable to the problem of inflated scores (e.g., Koretz, forthcoming). Evidence on this point remains scarce.

The findings reported here suggesting the potential for inflated score gains on KIRIS underscore the importance of validating KIRIS gains. Teachers' opinions on this question are made more credible by the finding that initial gains on KIRIS were not mirrored in scores on the National Assessment of Educational Progress or the American College Testing assessments (Hambleton et al., 1995). Nonetheless, survey data of this sort can only raise a warning flag, and additional research is needed to ascertain the validity of gains on KIRIS.

In evaluating this question, it is essential to distinguish between score gains during the initial years of a testing program and thereafter. Sizable score gains caused by increased familiarity are common during the first years of testing programs, and these gains need not be entirely misleading or undesirable. The issue changes complexion, however, after a few years.

First, whether initial gains caused by familiarization should be considered "real" or "inflated" depends on the circumstances and the inferences the test is used to support. For example, suppose a decision is made that certain types of complex problem-solving are

important outcomes of mathematics instruction and should be given much more weight on a new test than on an old one. If students and teachers learn what types of problem-solving are valued as outcomes as a result of familiarity with the test and students score higher as a result, that increase in scores would be considered a real gain. In contrast, if scores improve because students and teachers learn which of several alternative formats or which subsets of valued outcomes are likely to be emphasized and focus their efforts on those particular things at the cost of reducing emphasis on other important formats or sets of outcomes, the meaningfulness of the resulting gains is questionable. It was for this reason that our surveys asked about the impact of improvements in the knowledge and skills emphasized by KIRIS (as well as broad improvements in knowledge and skills). The fact that relatively few educators reported that even those more focused improvements contributed a great deal to score gains in their schools is a warning flag that suggests a need for further research.

Second, score gains attributable to familiarity might represent an increase in validity even when they do not represent commensurate gains in mastery. Because KIRIS represented such a large change in the content and format of KDE's assessments, it is plausible that scores in the first years of the assessment were misleadingly low. That is, scores may be lower than students' mastery warranted—for example, if their performance was impeded by unfamiliar task formats. As students become more familiar with those aspects of the assessment, scores would be expected to rise and to become a more accurate indicator of what students know.[26]

However, familiarity may also enable teachers to engage in forms of test preparation that can inflate scores—for example, tailoring instruction so closely to the details of the assessment that the resulting gains are too specific to the test to represent meaningful improvements in the skills the test is supposed to measure. The potential for this is greater when stakes are high because the incentives to raise scores per se are stronger.

The question for both policy and research is therefore more difficult than whether some share of KIRIS gains can be attributed to familiarity or test preparation. Rather, the key questions are *what share* of gains represents meaningful improvements and, conversely, what share is test-specific. In this regard, it is important to recall that although teachers were more likely to point to test preparation and familiarity as having contributed a great deal to their schools' gains on KIRIS, some did report that improved knowledge and skills had contributed a great deal, and most said that such improvements contributed at least a moderate amount. In addition, it would be valuable to obtain information on variations in test preparation and score inflation. For example, if appreciable score inflation is present, it would be important to know whether it is more severe in certain types of schools (e.g., those with initially low achievement), whether scores are more accurate in some subjects than others, and which forms of instruction and test preparation are related to score inflation.[27]

---

[26]Some workers in educational measurement use the term "familiarity" primarily in this way, that is, to refer to students' learning about construct-irrelevant aspects of a test. Therefore, they use "gains caused by familiarity" to refer primarily to the increase in validity that familiarity can cause.

[27]Previous research suggests that inflation of scores may be more prevalent in mathematics than in reading (Koretz et al., 1991).

Moreover, the validity of gains over the coming few years is a more pressing question than that of the initial gains about which we questioned Kentucky educators. The initial familiarization, which obscures the meaning of score gains, should be largely completed by this time. Moreover, when the initial effects of familiarity have run their course, some teachers may increase their reliance on narrow forms of test preparation in an effort to maintain the gains they experienced in the first years of the assessment's use.

Thus these findings appear to warrant further research on the validity of score gains and its correlates. If research revealed inflated scores, one could design program modifications in an effort to make undesirable test preparation less common or less effective. These might include, for example, changes in the specificity of curriculum frameworks, new guidelines distinguishing between appropriate and inappropriate test preparation, and perhaps modifications to the assessment itself, such as changes in content coverage or in the sampling of task types.

Portfolios: Impact and Validity

Teachers reported that portfolios had positive effects on teaching, causing them to be more innovative in terms of planning and instruction. They also indicated that portfolios made it more difficult to cover the curriculum, and they shifted emphasis away from mechanics and computation as a result. It appears that both of these changes are consistent with the goals of KERA, and they represent a success for the reform. However, curricular changes need to be monitored over time to make sure they remain consonant with KERA and that teachers do not go too far in emphasizing the "new" skills over the old.

Researchers who study assessment reforms have pointed out a fundamental tension between assessment as an inducement to instructional reform and assessment as a measurement tool (Koretz et al., 1994a). This tension is quite clear in the case of the Kentucky portfolios. On the one hand, teachers individualized their portfolio practices to reflect the needs of students. For example, they offered individualized help to students completing portfolio entries. For some students this meant helping them understand the nature of the task or problem; for others this meant helping them express their ideas clearly. Teachers also customized the program to suit their own styles and expectations by adopting different procedures regarding the completion of portfolio entries (e.g., number of revisions). By individualizing and customizing, teachers better integrated the portfolios into their instructional program.

On the other hand, standardization is important if assessment is going to be used as the basis for comparisons between schools, particularly if the results of the comparisons are to be used as a basis for accountability. If the products to be scored are not produced under similar conditions, then the scores that are assigned cannot be fairly compared. In the case of the Kentucky portfolios, the variation between classrooms and schools in teachers' portfolio practices (including the number of times pieces were revised, the amount of time devoted to a typical piece, the level of difficulty of pieces in students' portfolios, and the amount and type of assistance provided by teachers) potentially undercuts the validity of comparisons among schools, including comparisons among schools in growth over time.

Thus, the same features that make portfolios instructionally desirable threaten their use for accountability. It is an open question whether particular variations in portfolio practices in Kentucky (e.g., more time for revision, greater teacher assistance, more attention to scoring criteria) substantially influenced scores, but this is an issue that warrants further study.

Other Issues of Validity

Several of the findings reported here point to a need for additional studies of the validity of KIRIS, apart from the essential validity question of possible score inflation. Respondents raised concerns about the characteristics of the assessment itself that could affect the validity of many important inferences. Examples include concerns about the inconsistency of content representation and about tasks that may be developmentally inappropriate. Teachers also raised concerns pertaining to the assessment's use. For example, many pointed to possible distortions of inferences about school performance and improvement stemming from irrelevant factors such as students' backgrounds and transience. (These distortions, if they are present, would be similarly germane if KIRIS were replaced with a traditional, standardized, multiple-choice test used for similar purposes.) Such responses also raise concerns about "consequential validity"—that is, the possibility that the changes in education induced by the program may be less consistently positive than intended.

Next Steps

Both KIRIS and the broader KERA reform of which it is a key element are viewed nationwide as pathbreaking attempts to use innovative assessments as the engine of standards-based reform. The responses of educators to these surveys suggest that the program is meeting with some important initial successes; for example, educators perceive positive effects on instruction, and large numbers of them have come to accept and value innovative assessments. At the same time, however, these findings also suggest tensions, obstacles, and unintended negative effects. The results of these surveys should be useful both in planning modifications to the program and in charting additional investigations that could prove invaluable as the Kentucky Department of Education continues its efforts to improve the program.

The evaluation of KIRIS is likely to be a complex and long-term process. Validation requires numerous types of information, particularly in the case of assessments such as KIRIS that use innovative performance-assessment formats, serve multiple functions, and are designed to change instruction (see, for example, Linn, Baker, & Dunbar, 1991; Messick, 1995). Traditional forms of validity evidence—for example, convergent and divergent evidence about the relationships between KIRIS scores and other measures—will be essential, but it may also be important to use less commonplace techniques. For example, protocol analysis may be helpful for ascertaining the developmental appropriateness of tasks. Ascertaining the validity of hybrid tasks performed partly in groups and partly alone may also require innovative techniques (see, for example, Webb, 1993). Validation of gains may require extensive data collection, such as the administration of audit tests. Validation might

also be facilitated by ascertaining the relative difficulty of new and old KIRIS items in the absence of specific test preparation—for example, by conducting equating studies in which both new and old items are administered to students outside of Kentucky. Similarly, evaluating the effects of the program on education will likely require diverse information, including surveys and more intensive case studies.

Moreover, KIRIS is still a young program, and both the validity of its scores and its effects on education are likely to change as it matures. Ongoing research and evaluation will be needed to track the effects of these changes.

REFERENCES

Advanced Systems in Measurement and Evaluation (1993). *KIRIS 1991–92 technical report.* Frankfort, KY: Kentucky Department of Education.

Afflerbach, P., Guthrie, J., Schafer, W., & Almasi, J. (1994). *Barriers to the implementation of a statewide performance program: School personnel perspectives.* Paper presented at the annual meeting of the American Educational Research Association.

David, J. L. (1994). School-based decision making: Kentucky's test of decentralization. *Phi Delta Kappan, 75(9)* 706-712.

Druker, S. L., & Shavelson, R. J. (1995). *Assessment reform from both sides of the fence: Reformers' expectations and teachers' reports of impacts on classroom practice.* Paper presented at the annual meeting of the American Educational Research Association.

Guthrie, J. T., Schafer, W. D., & Afflerbach, P., & Almasi, J. (1994). *District-level policies of reading instruction in Maryland and their relation to the state-wide performance assessment.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April.

Hambleton, R. K., Jaeger, R. M., Koretz, D., Linn, R. L., Millman, J., & Phillips, S. E. (1995). *Review of the measurement quality of the Kentucky instructional results information system, 1991–1994.* Frankfort, KY: Office of Education Accountability, Kentucky General Assembly.

Kentucky Department of Education (1993). *Kentucky instructional results information system: 1991–92 technical report.* Frankfort, KY: Kentucky Department of Education.

Kentucky Department of Education (1994). *Kentucky instructional results information system: 1992–93 technical report.* Frankfort, KY: Kentucky Department of Education.

Koretz, D. (forthcoming). Using student assessments for educational accountability. In R. Hanushek (Ed.), *Improving the performance of America's schools.* Washington, DC: National Academy Press.

Koretz, D. M., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991). The effects of high-stakes testing: Preliminary evidence about generalization across conventional tests. In R. L. Linn (Chair), *The effects of high stakes testing,* symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago, IL, April.

Koretz, D., Mitchell, K., Barron, S. I., & Keith, S. (1996). *Perceived effects of the Maryland school performance assessment program.* Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Assessment (University of California at Los Angeles).

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994a). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice, 13(3),* Fall, 5–16.

BEST COPY AVAILABLE

Koretz, D., Stecher, B., Klein, S., McCaffrey, D., & Diebert, E. (1994b). *Can portfolios assess student performance and influence instruction? The 1991–92 Vermont experience.* RP-259. Santa Monica, CA: RAND.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20(8)*, 15-21.

McLaughlin, M. W. (1990). The RAND change agent study revisited: Macro perspectives and macro realities. *Educational Researcher, 19(9)*, 11–16.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14(4)*, 5–8.

Public Law 103-382 (1994). *Improving America's schools act of 1994.* 108 Stat. 3518, October 20.

Shepard, L. A., & Dougherty, K. C. (1991). Effects of high-stakes testing on instruction. In R. L. Linn (Chair), *The effects of high stakes testing,* symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago, IL, April.

Stecher, B. M., & Mitchell, K. J. (1995). *Portfolio driven reform; Vermont teachers' understanding of mathematical problem-solving.* Los Angeles, CA: CRESST/UCLA (CSE Technical Report 400).

Webb, N. W. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. *Educational Assessment, 1(2)*, 131–152.

MR-792-PCT/FF

ISBN 0-8330-2435-3

84